Candidate Enhancer and Transcription Dynamics in Early *Xenopus tropicalis* Development

Rosa Gomes Faria

University College London
and
The Francis Crick Institute
MRC – National Institute for Medical Research

A thesis submitted for the degree of
Doctor of Philosophy
University College London

December 2017

Declaration

I, Rosa Gomes Faria, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____

Rosa Gomes Faria

Abstract

A previous study from the Gilchrist lab measured transcription in the *Xenopus tropicalis* embryo for the first 66 hours of development at a high temporal resolution. This data showed that early transcription in *Xenopus tropicalis* is a very dynamic process, indicating that the mechanisms regulating it should also present dynamic behaviours.

The aim of this project was to understand the dynamics of enhancer usage and investigate whether this correlates with gene transcription, using p300 as an enhancer proxy and *Xenopus tropicalis* as a model. A 12.5-hour p300 ChIP-seq high-resolution time series, covering the end of the blastula stages, gastrulation and most of neurulation, was generated. ChIP-seq time series data analysis was optimised, including through the development of a normalisation method which allows for varying levels of background reads in different ChIP-seq samples. A dataset of 9,807 candidate enhancers and their respective usage dynamics was generated, a potentially useful tool for the *Xenopus* community. Furthermore, I showed that p300 binding dynamics at promoters and nearby candidate enhancers correlate well, reinforcing the enhancer-promoter loop model for transcription regulation. Additionally, I showed that p300 binding dynamics at promoters correlate with gene transcription, suggesting that the loop is maintained for the duration of transcription. I used both results to create a method to predict candidate enhancer-gene pairs and, with the addition of differential DNA motif analysis, to predict candidate target genes for some well-known transcription factors. The data generated in this project helped shed light on transcriptional regulation and led to the development of some useful tools both for *Xenopus* and transcription researchers.

For grandad

Obrigada por existires

Acknowledgements

I would like to thank several people without whom this thesis would not have been possible.

Firstly I would like to thank Mike Gilchrist, for giving me the opportunity to do this PhD, for his guidance and advice. I would also like to thank Jim Smith for taking me into his lab in the final months of my PhD and, together with Peter Thorpe and Jean-Paul Vincent, for all the help and comments as part of my Thesis Committee. Thanks to Brook and Elena for all the frog and lab advice and help, to Ian and Ilya who had to suffer with my endless computational questions and the biggest thank you to Nick, for inspiring me and always being there to help me. You all made the Gilchrist lab a great group to work with. Thanks to the animal and the sequencing facility members for all the technical support and thanks to every past and current member of the Elgar and Smith lab for all the help, as well as availability for parties! A special thank you to Bathilde and Christina for listening to my constant whining and for all the fun in the 4NE corner of the Crick. And of course, to Rita, my fellow Portuguese, doing our PhDs at the same time was a gift and you were always my go-to person during these four years, for science advice and gossip ("*o que é o vento?*"). Thanks to all the fellow Portuguese people for our Monday lunches, making me feel closer to home.

This PhD was done as part of the DevCom Marie Curie Initial Training Network. I would first like to thank the European Union for funding me, with the hope that many more people after me may get the same privilege. I would like to thank everyone that was part of it, it was the best opportunity of my life. A special thank you to Gert Jan Veenstra and José Luis Goméz-Skarmeta for all the advice and collaboration. Thanks to Simon van Heeringen for the long conversations and

email chains, your help and endless patience was essential for this PhD. Thanks to all the students for making all the courses and conferences so much fun.

I would also like to thank all my friends and family, unfortunately I cannot mention everyone, but you were all essential and a bit of this thesis is also yours. Thanks to all the PG people, especially my three “Pauls”, you kept me sane in these past five months. Thanks to Joe and Meena for all the party nights but also serious and supportive conversations. Thanks to Carolina, Filipa and Rita for our travels and skype calls. Thanks to my two “persons”, Diane and Pedro, I miss you so much but know you are always there for me. Thanks to Rita (Táta) for the advice on how to make pretty images and for pushing me to the finish line. Thanks to Pádua for arranging my post-PhD holidays. Thanks to Jane, Nigel, Holly and Emily for taking me into their family and always making me feel at home. Thanks to Sofia, Leonor and Rita for filling my heart with pure love. Thanks to Inês and Ana for always supporting and believing in me, “*Ró consé!*”. Pai and Mãe, no words can ever thank you for all you have done for me, I hope I have made you proud. And to Tom, for being there every day, through the highs and lows that a PhD brings. Thanks for making me look forward to the future.

Table of Contents

Table of Figures

List of tables

Abbreviations

3C	Chromosome conformation capture
ac	Acetylation
ar	ADP ribosylation
ATAC-seq	Assay for Transposase-Accessible Chromatin with sequencing
BMP	Bone morphogenetic protein
bp	Base pair
CAGE	Cap analysis of gene expression
CBP	CREB-binding protein
CE	Candidate enhancer
ChIP	Chromatin Immunoprecipitation
CTCF	CCCTC-binding protein
DHS	DNaseI hypersensitive
DNA	Deoxyribonucleic acid
E	Glutamic acid
ED	Euclidean Distance
ENCODE	ENCyclopedia Of DNA Elements
EST	Expressed sequence tag
Fast	Forkhead activin signal transducer
Foxh	Forkhead box H
Grhl	Grainyhead-like
GWAS	Genome-wide association studies
H3	Histone 3
H4	Histone 4
K	Lysine
Hi-C	Chromosome conformation capture with high-throughput sequencing
hiNOS	Human inducible nitric oxide synthase
HMG	High mobility group
HP1	Heterochromatin Protein 1
hpf	Hours post fertilisation
Lef	Lymphoid enhancer factor
Lmo	LIM domain only
MBT	Mid-blastula transition

Me1/2/3	Mono, di, trimethylation
mRNA	Messenger RNA
MZT	Maternal-to-zygotic transition
PIC	Preinitiation complex
ph	Phosphorylation
PRC2	Polycomb repressive complex 2
R	Arginine
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
S	Serines
SED	Sum of Euclidean Distances
shh	*sonic hedgehog*
SNP	Single-nucleotide polymorphism
Sox	SRY-box
su	Sumoylation
T	Threonines
TADs	Topologically associated domains
TAF1	TBP-associated factor 1
Tcf	T-cell factor
TSS	Transcription start site
ub	Ubiquitylation

Chapter 1. Introduction

1.1 Transcription and its Regulation

All cells of an organism contain virtually the same DNA sequence so, in order for cells to differentiate and perform the stereotypical functions of their specific cell type, spatio-temporal patterns of gene expression need to be tightly regulated. This is primarily achieved through the regulation of the production of transcripts. Transcriptional regulation is achieved through highly complex regulatory networks that exert control through several different mechanisms, such as modulation of chromatin structure, protein availability, transcription initiation, elongation and mRNA splicing (reviewed in Maston et al., 2006).

A key step in transcriptional regulation is the modulation of the initiation phase, which involves different DNA elements, epigenetic modifications and the recruitment of general and sequence-specific transcription factors to their target sites on DNA. The DNA sequences involved are the promoter, adjacent to the transcription start site (TSS), and distal elements such as enhancers and insulators (Figure 1) (reviewed in Maston et al., 2006, Spitz and Furlong, 2012).

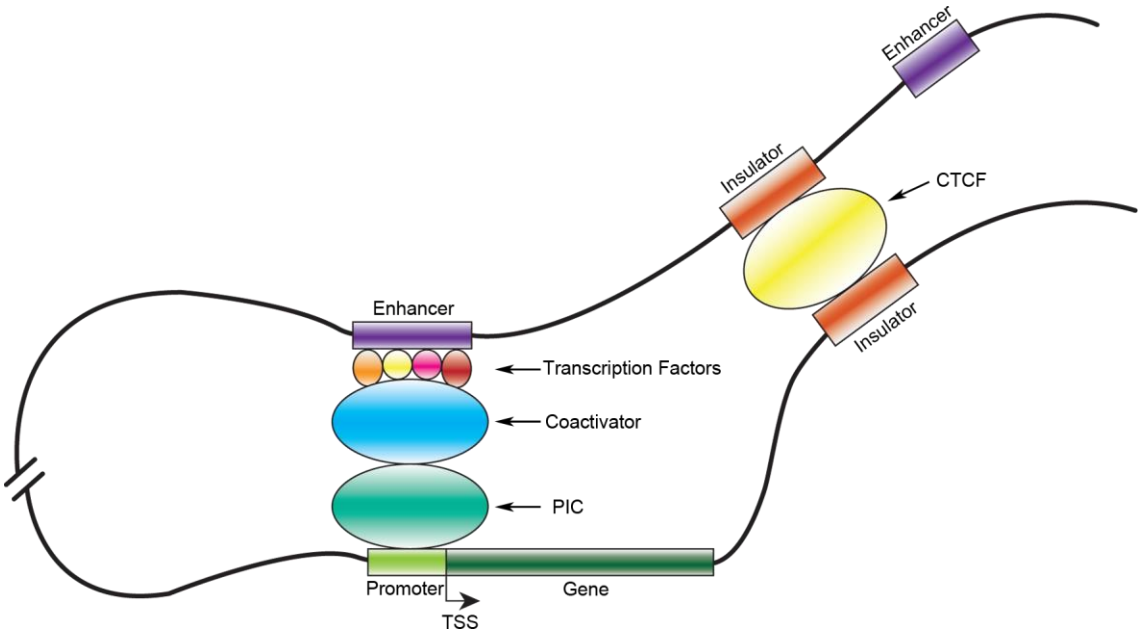


Figure 1 - Model of eukaryotic gene regulation.

Sequence-specific transcription factors bind to their motifs in enhancer regions (purple) and recruit coactivators (blue), such as p300, which mediate enhancer-promoter loop formation, histone remodelling and/or RNAPII recruitment. The preinitiation complex (PIC) (green) assemble in the promoter region (light green), containing the TSS, where transcription starts. Insulators (orange) are bound by CTCF (yellow) and form a loop to prevent incorrect enhancer-gene pairings.

The promoter is where the preinitiation complex (PIC) is assembled before initiating transcription. The PIC is formed by RNAPII, general transcription factors, such as the TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, and the Mediator (reviewed in Maston et al., 2006). The Mediator, a large protein complex, interacts with sequence specific transcription factors, transmitting its signals to RNAPII, modulating its function, e.g. by phosphorylating it, leading to the release from the PIC. The Mediator is involved in transcriptional regulation by altering chromatin organisation and modulating enhancer-promoter looping and transcription initiation and elongation (reviewed in Allen and Taatjes, 2015).

Transcription initiation can be modulated by the combinatorial binding of sequence-specific transcription factors to enhancer sequences (reviewed in

Kadonaga, 2004). Coactivators, in turn, modulate transcription by binding to sequence-specific transcription factors and to the PIC, creating a DNA-loop structure. Coactivators can also promote PIC assembly, modify the chromatin structure and/or mediate DNA unwinding (reviewed in Spiegelman and Heinrich, 2004, Maston et al., 2006).

The ENCODE (ENCyclopedia Of DNA Elements) Project set out to identify and explain how all the different types of functional sequences act together and lead to all the different cellular processes (Encode, 2004). They first analysed 1% of the human genome and detected that most base pairs are transcribed; identified previously unknown TSSs and non-coding transcripts; and identified several chromatin modifications which are characteristic of different types of functional elements (Encode et al., 2007), which will be explored below (1.1.4.2 – Histone modifications). The Encode Project has helped further the knowledge of how transcription is regulated and their studies will be mentioned throughout this introduction.

1.1.1 Enhancers

Banerji and colleagues first used the term “enhancer” to describe how fragments of SV40 viral DNA were able to increase transcription of the transfected rabbit β-globin gene in HeLa cells by 200 times (Banerji et al., 1981). The authors also showed how enhancers are effective in both orientations and both upstream and downstream of the gene promoter. Two years later the same team identified the first eukaryotic enhancer, for the mouse β-globin gene (Banerji et al., 1983).

Usually enhancers act on the closest gene (de Villiers et al., 1983, Wasylyk et al., 1983, Dillon et al., 1997) however there are multiple exceptions, with enhancers acting on genes further away (de Laat and Grosveld, 2003, Spitz et al., 2003, Shi et al., 2013), as in the classical example of the sonic hedgehog (*shh*) enhancer, which is located 1 Mb away from the gene’s TSS, inside an intron of another gene (Lettice et al., 2003). A recent study reported that 10% of promoter interactions were with regions more than 1 Mb away and some interactions are even interchromosomal (Javierre et al., 2016). Not all of these interactions are expected to be functional, however, at least some are likely to be, which shows that gene regulation can be achieved over extremely long distances.

Enhancers usually contain the binding site sequences for several transcription factors (Zeitlinger et al., 2003, Mullen et al., 2011, Trompouki et al., 2011). Genes may also have multiple enhancers, each being active at different developmental stages and/or at different cell types, or they may act synergistically to drive gene expression (reviewed in Maston et al., 2006, Shlyueva et al., 2014).

O’Kane and Gehring developed a method to systematically identify enhancer regions in *Drosophila,* the enhancer trap. This method uses transposable elements with a reporter gene, which randomly integrates in the genome, and the

reporter gene's expression is then analysed and the flanking regions sequenced to identify the regulatory region (O'Kane and Gehring, 1987). Ruf and colleagues developed a similar method to probe regulatory elements in mouse embryos, also by using a transposon system (Ruf et al., 2011). A reporter gene, with a minimal promoter unable to drive gene expression, is inserted in random regions and their expression, driven by endogenous enhancers, is analysed. They showed that most insertions had very restricted tissue expression, indicating that the majority of enhancers regulate gene expression in a tissue-specific way; even genes ubiquitously expressed will likely have different enhancers driving expression in different tissues.

Enhancer regions contain clusters of motif sequences, usually 6 to 12 bp long, to which transcription factors bind, usually in a combinatorial and synergistic way (reviewed in Kadonaga, 2004, Shlyueva et al., 2014). However, the majority of transcription factor binding events in the genome do not seem to lead to expression of neighbouring genes; in yeast it has been estimated that only 50% of binding is functional (Gao et al., 2004, Ucar et al., 2009), with the number for higher eukaryotes being reduced to around 20% (Vokes et al., 2008, Krejci et al., 2009). There may be several reasons behind the low percentage of transcription factor binding events leading to gene transcription: transcription factor redundancy; random binding events in areas of open chromatin; the need for additional cofactors; or the binding may be achieving other functions other than directly regulating transcription, such as chromatin remodelling (reviewed in Spitz and Furlong, 2012).

1.1.1.1 Enhancers and disease

Studying enhancer regions and how they are involved in transcriptional regulation is essential to understanding development, but also because mutations in enhancer regions have been identified as the cause of several disorders in humans. Mutations in a *shh* enhancer (1 Mb away from the gene's TSS) cause preaxial polydactyly (Lettice et al., 2003); individuals with a common variant in an enhancer for the RET gene have 20-fold increased risk for Hirschsprung's disease (Emison et al., 2005); and a single-nucleotide polymorphism (SNP) on an enhancer 300 kb upstream of the *MYC* oncogene leads to increased cancer risk (Sur et al., 2012), among many other examples (reviewed in Iyer et al., 2004).

More than one third of SNPs identified in genome-wide association studies (GWAS) do not overlap known exons, suggesting that a high proportion of altered phenotypes are caused by mutations in non-coding sequences (Visel et al., 2009b). The ENCODE Project reported an even higher value in their dataset of functional human elements, with 88% of disease associated SNPs being present in non-coding regions of the genome (Encode, 2012).

1.1.2 Insulators

Insulators, first described in the early 90s in *D. melanogaster* (Holdridge and Dorsett, 1991, Kellum and Schedl, 1991, Kellum and Schedl, 1992), are DNA sequences that prevent contacts between enhancers and promoters on opposite sides of the insulator, inhibiting inappropriate gene activation. Insulators can also separate regions with different levels of chromatin condensation (Donze et al., 1999, Prioleau et al., 1999) (Insulators reviewed in West et al., 2002). It was later found that CCCTC-binding factor (CTCF) binds insulators (Chung et al., 1993, Bell et al., 1999) and that CTCF sites can contact each other creating a chromatin loop (Splinter et al., 2006). Insulators are outside of the scope of this thesis, but for an excellent review, I would direct the reader to (Gaszner and Felsenfeld, 2006).

1.1.3 Enhancer-promoter looping

Enhancers and promoters can physically interact through chromatin looping, first shown in the repression of bacterial genes (Ptashne, 1986) and later in the mice β-globin locus (Carter et al., 2002, Tolhuis et al., 2002). Palstra and colleagues then showed that, for that same locus, during development, the topology of the loop changes depending on which genes are being expressed (Palstra et al., 2003). Several studies have identified PIC proteins, such as Mediator subunits and general transcription factors, at enhancers (Lin et al., 2013, Zhou et al., 2013), suggesting that the enhancer and the promoter are at very close proximity during transcription initiation (reviewed in Levine et al., 2014). The enhancer-promoter loop structure and how it is formed is still not fully understood, however there is evidence that the two regions are linked by Mediator-cohesin

protein complexes, which form rings around the DNA (Kagey et al., 2010, Phillips-Cremins et al., 2013).

About half of enhancer regions interact with multiple promoters and most promoters interact with multiple enhancers (Thurman et al., 2012, Javierre et al., 2016). It was proposed that enhancers contact multiple promoters in the same cell through a “flip-flop” mechanism, with the enhancer looping to one promoter and then another (Wijgerde et al., 1995). However, Fukaya and colleagues have recently shown that an enhancer can contact and activate multiple gene promoters simultaneously, in the same cell, instead of only one at a time (Fukaya et al., 2016).

In 2002, Dekker and colleagues developed a technique that allowed the identification of spatial colocalisation between two chromosomally distant sequences – Chromosome conformation capture, or 3C (Dekker et al., 2002). Several improvements have been made to this technique, leading to the development of 3C-based methods like 4C, 5C and Hi-C, allowing the probing of more and more genomic regions (reviewed in de Wit and de Laat, 2012, Denker and de Laat, 2016). These advances allowed the discovery that chromosomes are divided into topologically associated domains (TADs), with DNA sequences being more likely to contact sequences inside the same TAD. These domains tend to be maintained between different tissues (Dixon et al., 2012, Nora et al., 2012, Sexton et al., 2012), however, these boundaries are not absolute, and it has been reported that about 30% of promoter-enhancer interactions are across TAD boundaries (Javierre et al., 2016).

TAD boundaries are enriched in CTCF (Dixon et al., 2012) and when disrupted, gene promoters start interacting with different enhancers, which can lead to altered gene expression and consequent disorders (Lupianez et al., 2015).

Ghavi-Helm and colleagues performed 4C-seq experiments in Drosophila embryos at two developmental stages and in mesoderm cells versus whole embryos (Ghavi-Helm et al., 2014). They reported that most enhancer-promoter loops are stable and do not change between different stages and tissues, even when expression of genes within those TADs varies. The authors suggest that when the enhancer loops to interact with the promoter, RNAPII is recruited but remains paused; later on, at the time of gene activation, transcription factors and/or other enhancer regions may be recruited leading to the release of RNAPII pausing, activating transcription. Despite this study suggesting enhancer-promoter loops are maintained, several other studies report changes in looping and enhancer states during development (Tolhuis et al., 2002, Simonis et al., 2006, Heintzman et al., 2009). Ji and colleagues have recently reported that TADs are maintained in different cell types, but the contacts inside them vary depending on the cell type and the genes expressed (Ji et al., 2016). 3C-based methods are based on average interactions in thousands or millions of cells; more stable and long-lasting interactions will be identified more often, which may bias the data and underestimate the number of fast and dynamic loop structures and their importance in transcriptional regulation.

Further evidence that enhancer-promoter looping is a dynamic process that varies depending on whether a gene is being expressed was given by a recent study by Javierre and colleagues (Javierre et al., 2016). They performed promoter capture Hi-C (a 3C-based method applied genome-wide), in which Hi-C fragments are pulled down if they contain promoter regions. This study showed that promoter-enhancer interactions are highly specific, depending on the celltype and on the gene's activation state.

1.1.4 Chromatin structure

The chromatin structure surrounding promoters and enhancers is also essential for transcriptional regulation. It can facilitate or preclude access to DNA binding motifs, modulating where the different proteins involved are able to bind.

DNA is packaged in the cell in the form of chromatin, a DNA-protein complex. Its basic unit is the nucleosome consisting of 147 bp of DNA wrapped around an octamer of histones – two H2A, two H2B, two H3 and two H4 (reviewed in Kouzarides, 2007).

1.1.4.1 Accessible chromatin regions

Active enhancers, as well as promoters and other regulatory regions, are nucleosome-depleted, in order to increase DNA accessibility. Wu first described DNase I hypersensitivity in *D. melanogaster* gene promoters and recognised that those areas would allow for easier protein binding (Wu, 1980). DNase I hypersensitive (DHS) sites can now be determined genome-wide by DNase I-seq; this method uses DNases, nucleases which preferentially fragment areas of the chromatin which are less condensed, and the fragments are then sequenced and mapped to the genome (Boyle et al., 2008, Song and Crawford, 2010).

The ENCODE Project showed that TSS sequences have high DNAse I hypersensitivity and that 98.5% of transcription factor bindings in the genome occur within DHS sites (Encode et al., 2007, Encode, 2012). Lu and colleagues performed DNAse I-seq and showed that in mouse embryo development the genome is increasingly more accessible, with promoter regions being accessible before enhancers (Lu et al., 2016).

Although DNAse I-seq has allowed us to gain insights into the chromatin structure, it requires a very high number of cells and the optimization of the experimental conditions is not always straightforward. To overcome these limitations, Buenrostro and colleagues developed a technique called Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq) (Buenrostro et al., 2013). ATAC-seq is able to detect open chromatin in as few as 500 cells, through the use of Tn5 transposase and sequencing adaptors. In one single step the Tn5 fragments and tags the genome with adaptors, being more likely to do this in regions of open chromatin. These fragments are then amplified and sequenced (Buenrostro et al., 2013, Buenrostro et al., 2015). The authors also reasoned that DNA bound by other proteins, such as transcription factors, would be transposed less frequently, making it possible to detect discrete footprints (Buenrostro et al., 2013). This method is now widely used to detect open chromatin areas and DNA binding protein footprints (Davie et al., 2015, Wu et al., 2016, Jorstad et al., 2017, Younger and Rinn, 2017).

1.1.4.2 Histone modifications

Histones have N-terminal tails that can be modified leading to alterations in the nucleosome structure and, consequently, in DNA accessibility. There are several different types of chemical modifications, namely, acetylation (ac), methylation (me), ubiquitylation (ub) and sumoylation (su) of lysines (K), methylation of arginines (R), phosphorylation (ph) of serines (S) and threonines (T), ADP ribosylation (ar) of glutamic acid (E), deamination (conversion of arginine to citruline) and proline isomerisation. Methylation of lysines can be mono-, di- or trimethylation (me1, me2, me3) (reviewed in Kouzarides, 2007). Dozens of enzymes that mediate these modifications have been identified, however this is out of the scope of this thesis, Kouzarides, 2007 is an excellent review of all the identified enzymes and the mechanisms involved.

There are dozens of different histone modifications, here I will review the ones more relevant in transcriptional regulation.

1.1.4.2.1 H3K4me

H3K4me3 (histone 3, lysine 4 trimethylated) was shown to mark active promoters (Santos-Rosa et al., 2002), while H3K4me2 seemed to be present on both active and inactive genes (Schneider et al., 2004). In a study of 1% of the human genome, all three methylation forms of H3K4 were shown to be highly enriched in TSSs of active genes (Koch et al., 2007). H3K4me1, but not H3K4me3, is also enriched in enhancer regions, thus enhancers can be identified by a higher ratio of the monomethylated vs trimethylated modification, compared to promoter regions (Encode et al., 2007, Heintzman et al., 2007). Recently, Rickels and

colleagues reported that *Drosophila* embryos with reduced H3K4me1 develop normally and that gene expression is only minimally affected by the presence of H3K4me2/3 at enhancers, instead of H3K4me1. The authors suggest that H3K4me1 is only involved in the fine-tuning of transcription in specific situations, such as stress response (Rickels et al., 2017).

1.1.4.2.2 H3K36me

Strahl and colleagues reported that H3K36me at promoters is correlated with gene repression (Strahl et al., 2002). However, several groups have shown that H3K36me is correlated with active gene elongation, being deposited by the passage of the RNAPII (Krogan et al., 2003, Xiao et al., 2003, Mikkelsen et al., 2007). Inactive genes are associated with deacetylated histones and there is evidence that Rpd3, a histone deacetylase, is recruited by H3K36me in transcribed gene bodies and deacetylates the histones in order to repress incorrect initiation of transcription away from the TSS. This action is enriched at the 3' end of gene bodies, therefore the promoters maintain their acetylation, allowing gene transcription to be initiated at the correct site (Carrozza et al., 2005, Joshi and Struhl, 2005, Keogh et al., 2005).

1.1.4.2.3 H3K27me

H3K27me is present in the inactivated X chromosome (Plath et al., 2003, Silva et al., 2003) and is involved in gene repression by the Polycomb group (Kuzmichev et al., 2002). H3K27me3 is deposited by EZH2, part of the Polycomb repressive complex 2 (PRC2), which is known to repress developmental genes

(Boyer et al., 2006, Bracken et al., 2006). The monomethylated form of this histone tail is present in active genes, promoting their transcription; 70% of nucleosomes have the dymethylated form, inhibiting incorrect enhancer usage; while the trimethylated form is highly associated with gene repression (Barski et al., 2007, Ferrari et al., 2014). It has now been shown that all three methylation states of this histone tail are regulated by PRC2 (Ferrari et al., 2014).

1.1.4.2.4 H3K9me

H3K9me marks constitutive heterochromatin, due to its ability to recruit Heterochromatin Protein 1 (HP1) (Bannister et al., 2001, Lachner et al., 2001, Nakayama et al., 2001, Noma et al., 2001). H3K9me is also present in facultative heterochromatin in imprinted loci and is involved in X inactivation (Peters et al., 2002, Silva et al., 2003). H3K9me is also involved in silencing euchromatic regions, repressing gene transcription (Tachibana et al., 2002, Ayyanathan et al., 2003). Vakoc and colleagues detected that, surprisingly, H3K9 di- and trimethylation and HP1 are dynamically associated with active gene bodies, with both marks being quickly lost with the end of transcription (Vakoc et al., 2005). The authors blocked transcription elongation and H3K9me and HP1 levels decreased dramatically in gene bodies, while being maintained in heterochromatic regions, showing their deposition is correlated with transcriptional elongation. Wang and colleagues reported H3K9 di- and trimethylation in repressed areas, consistent with the early reports that this histone modification is present in inactive regions (Wang et al., 2008). H3K9me may activate gene transcription when present in gene bodies and be repressive when present at promoter regions (Kouzarides, 2007).

H3K4me and H3K9me3 are antagonistic, the former facilitates the action of p300, a transcriptional coactivator with histone acetyltransferase (HAT) activity, while the latter inhibits it (Wang et al., 2001, Nishioka et al., 2002a).

1.1.4.2.5 H3K20me

Nishioka and colleagues reported that H4K20me marks heterochromatin (Nishioka et al., 2002b). Later, the monomethylated version of that histone tail was found to be enriched in active genes (Talasz et al., 2005, Barski et al., 2007) and the trimethylated version in heterochromatic and repressed regions (Schotta et al., 2004, Wang et al., 2008).

1.1.4.2.6 Histone acetylation

Acetylation of lysines neutralises its charge, leading to an increase in chromatin flexibility and accessibility. Wang and colleagues studied the genomic distribution of 18 different histone acetylations and found they were all associated with gene activation, however some, including H3K18/27, are more prevalent in non-coding and non-promoter regions (Wang et al., 2008). As previously mentioned, enhancers are enriched with H3K4me1; Creyghton and colleagues found that H3K4me1 enhancers can either be active, inactive or poised and that the presence of H3K27ac is able to discriminate the active ones (Creyghton et al., 2010).

H3K9ac marks active regulatory elements, mainly promoters but also enhancers (Roh et al., 2005, Roh et al., 2007, Karmodiya et al., 2012). It has

recently been shown that H3K9ac is necessary for the release of paused RNAPII, stimulating elongation (Gates et al., 2017).

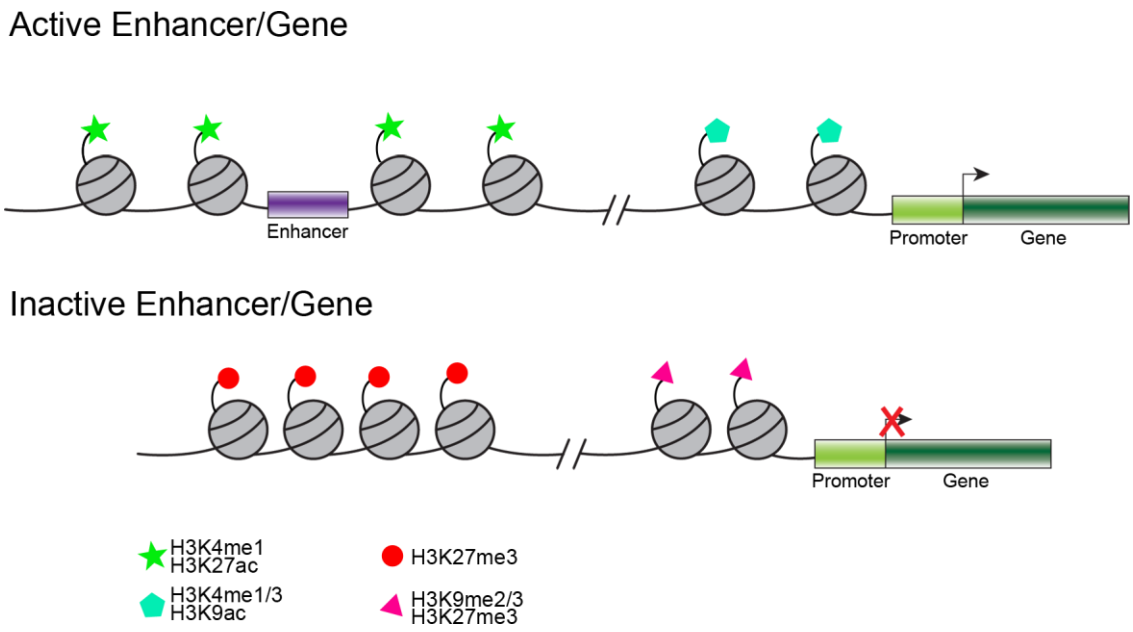


Figure 2 – Histone modifications and transcriptional regulation.

Histone modifications associated with active *vs* inactive gene transcription. H3K4me1 and H3K27ac are predictive of active enhancer regions (star); H3K4me1/3 and H3K9ac of active gene promoters (pentagon); H3K27me3 of repressed enhancers (circle); H3K9me2/3 and H3K27me3 of repressed gene promoters (triangles). (Adapted from Baldwin et al., 2013).

1.1.5 Enhancer RNAs

Kim and colleagues showed that 25% of neuronal enhancers, defined by the presence of CBP (a transcriptional coactivator) and H3K4me1, were transcribed by RNAPII (Kim et al., 2010). These *enhancer*-RNAs – or eRNAs - were found to be relatively short (<2kb), bi-directional, likely not polyadenylated and their expression levels correlated with the mRNA levels of nearby genes. The authors also suggested that eRNA synthesis may require an enhancer-promoter interaction; when the *Arc* gene (including its promoter) was deleted, RNAPII was still able to bind to the *Arc* enhancer but no eRNA was detected. Since then, many studies have reported these RNAs and have further increased our understanding of these molecules: eRNAs may be required for their associated gene's transcription (Mousavi et al., 2013, Iott et al., 2014, Schaukowitch et al., 2014); may be involved in regulating chromatin accessibility (Mousavi et al., 2013); in promoting and stabilising enhancer-promoter loops (Li et al., 2013) and in facilitating transcription elongation (Schaukowitch et al., 2014). eRNAs are also more likely to be produced in enhancers involved in promoter looping (Lin et al., 2012).

The knowledge that enhancers may be transcribed was used to predict enhancer regions, using CAGE (cap analysis of gene expression) which detects nascent transcripts (Andersson et al., 2014). This study predicted more than 40,000 enhancer regions based on the presence of bi-directional short eRNAs and reported that enhancer activity correlates highly with it. The authors also showed that 95% of eRNAs where unspliced, 80% were located in the nucleus, 90% were not polyadenylated and that they had a median length of 346 bp.

1.1.6 Transcriptional adaptor p300

p300 was first identified as an interaction partner of adenovirus E1A proteins, proteins which are able to regulate transcription of several genes (Harlow et al., 1986). Eckner and colleagues cloned and characterized it, describing it for the first time as a transcriptional adaptor (Eckner et al., 1994) and the same team later showed that it can stimulate transcription (Arany et al., 1995).

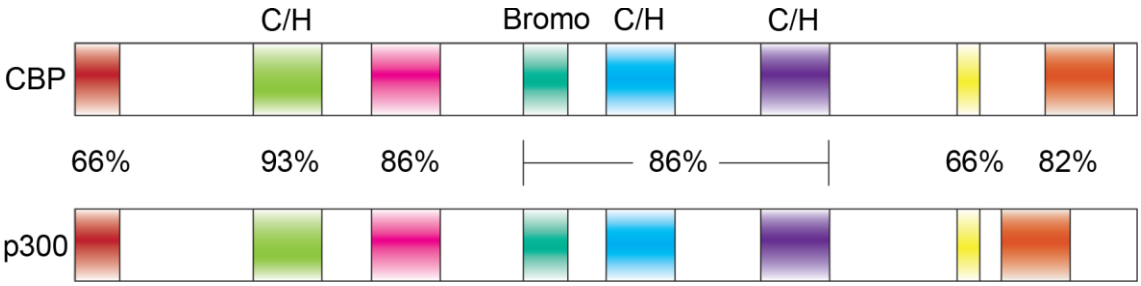


Figure 3 – CBP and p300 protein domains.
Percentages represent amino acid identity between the different domains, with the coloured areas representing areas with high identity. C/H – cysteine histidine rich zinc finger motifs; Bromo – bromodomain (Adapted from Giles et al., 1998).

CREB-binding protein (CBP) is a highly homologous protein to p300, with their known functional domains having more than 90% sequence identity. Their functional domains are conserved and they seem redundant in most functions (Arany et al., 1995, Lundblad et al., 1995) (reviewed in Roth et al., 2001). Figure 3 shows CBP and p300 protein domains and the regions of high identity. Both proteins are histone acetyltransferases (HATs), with their main targets being H3K18 and H3K27 (Jin et al., 2011), which are mainly present in non-coding, non-promoter regions, as mentioned above (1.1.4.2.6 – Histone acetylation).

p300 and CBP have multiple domains and numerous known interacting partners, including several transcription factors (Goodman and Smolik, 2000).

Being able to interact with a range of different proteins, p300 and CBP take part in several cellular processes, such as the previously mentioned transcriptional adaptor and histone acetyltransferase function, however, they are also involved in regulating the cell cycle (Yaciuk and Moran, 1991), growth (Howe et al., 1990, Wang et al., 1995), differentiation (Akimaru et al., 1997), apoptosis (Shikama et al., 1999) and DNA damage repair (Liu et al., 1999) (reviewed in Giles et al., 1998). In fibroblasts, p300 and/or CBP are required for the activation of some genes, while for others they are needed only to get maximal gene expression (Kasper et al., 2014).

1.1.6.1 p300 and transcriptional regulation

The role of p300 in transcriptional regulation is not yet fully understood. As mentioned above, p300 has affinity for several proteins; one example is the transcriptional coactivator ASC-2, which interacts with p300 as well as with sequence-specific transcription factors and with components of the promoter complex (Lee et al., 1999). It has also been recently shown that in human hepatocytes p300 binds RNAPII, at the hiNOS (human inducible nitric oxide synthase) promoter, and AP1, at the hiNOS enhancer, creating an enhancer-promoter loop (Guo et al., 2016). p300 knockdown led to the disappearance of this loop and decreased hiNOS expression by 50%. These interactions may help create the bridge between the sequence-specific transcription factors located at enhancers and the basal transcription machinery at promoters.

p300 may regulate transcription through a combination of the following mechanisms: the previously mentioned role in enhancer-promoter looping and its HAT activity; its ability to recruit (Janknecht and Hunter, 1996a, Janknecht and Hunter, 1996b, Kee et al., 1996, Cho et al., 1998) and acetylate RNAPII, which leads to its release from a paused state (Schroder et al., 2013); and its ability to acetylate and stimulate P-TEFb (a protein involved in transcription elongation) (Fu et al., 2007, Cho et al., 2009) (reviewed in Holmqvist and Mannervik, 2013). It has also been shown that p300 HAT activity alone can lead to gene activation; Hilton and colleagues used nuclease deficient Cas9 to guide p300's HAT domain to enhancer regions and saw an increased expression of the associated genes (Hilton et al., 2015). It has been shown that p300/CBP knockdown disrupts enhancer-

promoter looping, but not insulator looping (Kim and Kim, 2013), indicating that p300/CBP are involved in the former but not the latter type of looping.

In the same study in which eRNAs were first identified, the authors studied how CBP binding in primary neuronal cultures changed in response to high-level exposure to potassium chloride – which is known to activate calcium-dependent signalling pathways and change downstream gene expression (Kim et al., 2010). The number of CBP binding sites in the genome increased from 1,000 to 28,000 upon stimulation, suggesting that CBP is involved in the observed changes in gene expression. Furthermore, they analysed the binding of three transcription factors known to be involved in neuronal processes – CREB, SRF and NPAS4 – and found that they colocalise with CBP in a 200 bp region in the centre of the enhancer (Kim et al., 2010).

It has recently been shown that CBP binds eRNAs transcribed from CBP-bound enhancers, which stimulates its acetyltransferase activity, essential for gene regulation (Bose et al., 2017). The eRNAs bind to the CBP's HAT domain, which is highly conserved with p300's, thus, the same is expected to occur with the latter. The knockdown of CBP led to an 89% decrease in expression of the tested mRNAs and 50% of the tested eRNAs. This suggests that eRNAs are involved in stimulating CBP's HAT activity, but that CBP binding is needed for eRNA production, implying that the eRNAs are not involved in guiding or attracting CBP to the locus. Indeed, when eRNAs were depleted, CBP was still able to bind enhancers but enhancer and target promoter acetylation was significantly reduced.

1.1.6.2 p300/CBP and disease

Gayther and colleagues identified somatic mutations in the EP300 gene (which encodes p300) in several cancer samples and described it as a tumour-suppressor gene for the first time (Gayther et al., 2000). Since then, several studies have identified somatic mutations in the EP300 and CBP genes in cancer samples, which may be due to their interaction with proteins like p53, TGF-β and Rb, involved in tumorigenic pathways (reviewed in Iyer et al., 2004). The feasibility of targeting p300 for prostate cancer treatment was studied and showed promising results: p300 knockdown led to an increase in apoptosis and decrease in cell invasion (Santer et al., 2011).

Roelfsema and colleagues identified for the first time mutations in the EP300 gene in a congenital disease, Rubinstein-Taybi syndrome (Roelfsema et al., 2005). This syndrome causes mental and growth retardation and increased childhood tumour risk. This syndrome is more usually caused by mutations in the CBP gene, however, due to their high level of homology, the authors screened patients for mutations in the EP300 gene and found 3 inactivating mutations.

p300 or CBP homozygous and double heterozygous mouse knockouts are early embryonic lethal and p300 heterozygous knockouts had low survival rates; indicating that normal levels of these two proteins are essential for correct embryonic development (Yao et al., 1998, Oike et al., 1999a, Oike et al., 1999b).

1.1.6.3 p300 and enhancer prediction

In 2005, Wang and colleagues reported p300 binding to an androgen receptor gene enhancer, 4 kb away from the gene's TSS (Wang et al., 2005) and since then p300 has been widely used to predict enhancer regions.

Heintzman and colleagues performed Chromatin Immunoprecipitation (ChIP) analysis for p300, core histone H3, H4K5/8/12/16ac, H3K9/14ac, H3K4me1/2/3, RNAPII and TBP-associated factor 1 (TAF1), part of the basal transcriptional machinery, in 1% of the human genome selected by the ENCODE project. They found 70% of p300 binding sites were also DHS sites. More than 75% of p300 positive regions were located more than 2.5 kb away from known TSSs, 60% were conserved regions and 44% of sites coincided with previously described regulatory modules. p300 positive sites showed H3K4me1 enrichment, however this histone modification has a broader peak than p300. H3K4me3 was not found to be enriched at these sites. There was also a modest enrichment for RNAPII and TAF1 at enhancers, which may be an indication of a physical interaction with promoter regions (Heintzman et al., 2007).

Wang and colleagues compared p300 and CBP binding in human CD4+ T cells and showed that p300 is enriched in promoters and in candidate enhancers, with CBP having a very similar distribution (Wang et al., 2009).

Visel and colleagues predicted enhancer regions using p300 ChIP-seq data for forebrain, midbrain and limbs of mouse embryos. They then tested 86 of those regions for enhancer activity using transgenic mice, with 87% of sequences leading to reproducible expression patterns. This represented a 5 to 16-fold increased success rate when compared to studies which used sequence conservation to predict enhancers (Visel et al., 2009a).

In another study by Heintzman and colleagues, p300 and CTCF binding and histone modifications were analysed in 1% of the human genome, in five cell lines. They observed that promoters and insulators mostly have a constant chromatin state between cell types, unlike enhancers, whereas p300 binding and histone modifications change significantly depending on the cell type. About 80% of the identified enhancers were cell type specific, however, 85% of the active genes were common between the cell lines (Heintzman et al., 2009). This is consistent with what was reported by Ruf and colleagues (previously mentioned in section 1.1.1 – Enhancers), that even ubiquitously expressed genes are controlled by different enhancers in different cell types (Ruf et al., 2011). In the human genome, most of the ubiquitous DHS sites are at promoter or insulator regions, while cell type specific DHS sites are mainly in enhancer regions (Xi et al., 2007). All of these results show how dynamic enhancers' states are, particularly when compared to more stable elements, such as promoters and insulators. This indicates that enhancers are likely essential for cell type-specific gene expression.

In a study focusing on mouse embryonic heart tissue, more than 3,000 heart enhancers were predicted using p300 ChIP-seq data (Blow et al., 2010). The authors then tested 130 of the predicted sequences by transgenesis and 75% of those showed tissue-specific activity, a 29-fold increase in positive signals compared to studies using evolutionary conservation. The same team then predicted more than 6,000 heart enhancers in human fetal and adult heart tissue and 66% of the sequences tested by transgenesis drove expression in the heart. The two studies were compared and only 21% of fetal human heart enhancers overlapped with mouse heart enhancers, indicating a low evolutionary conservation of heart enhancers (May et al., 2011).

Attanasio and colleagues performed p300 ChIP-seq on facial tissue from mouse embryos in order to study craniofacial development and identified more than 4,000 enhancers. Transgenic mice assays were used to test 206 candidate regions and 60% showed activity in craniofacial tissues (Attanasio et al., 2013).

As mentioned above (1.1.4.2.6 – Histone acetylation), Creyghton and colleagues had shown that H3K27ac could distinguish active from poised/inactive enhancers (Creyghton et al., 2010). Rada-Iglesias and colleagues also found that p300-bound enhancers could be distinguished between active or poised based on the presence or absence of H3K27ac, respectively (Rada-Iglesias et al., 2011). They also tested non-conserved human embryonic stem cells' poised enhancers in zebrafish and they were able to drive expression in a time and cell type specific pattern.

Based on these studies, p300 binding would appear to be a good enhancer predictor, with 60-87% of p300 binding regions seeming to behave as enhancers in transgenic assays. Active enhancers can be predicted by the presence of p300, H3K4me1 and H3K27ac. Enhancer state seems to be particularly dynamic, compared to promoters and insulators, reinforcing the importance of these sequences for transcriptional regulation.

1.1.7 Sequence conservation

Sequence conservation is widely used to predict enhancer regions, however evidence is gathering for this not being the most reliable method. By comparing transcription binding sites in human and mouse hepatocytes, it was found that 41-89% of regions were species specific (Odom et al., 2007). In a study comparing human, mouse and dog CEBPA and HNF4A genomic distribution, only 10-22% of binding sites were conserved (Schmidt et al., 2010). It was also shown that the most commonly used metrics to detect evolutionary constraint were only able to detect 29-61% of known regulatory sequences (McGaughey et al., 2008, Bulger and Groudine, 2011) and this number may be an overestimate, due to many known regulatory regions having been identified based on sequence conservation.

As mentioned above (1.1.6.3 – p300 and enhancer prediction), studies which used p300 to identify enhancers regions found that only a small percentage of those were conserved amongst different species and that predicting enhancers by p300 binding increased the efficacy when compared to predicting enhancers only by sequence conservation (Visel et al., 2009b, Blow et al., 2010, May et al., 2011).

These findings are consistent with a model of enhancers being less conserved than promoters (Villar et al., 2015). This shows that enhancer prediction should not be performed solely by analysing sequence conservation and that analysing p300 binding and histone modifications can be more efficient.

1.1.8 Enhancer validation

The most common approach to test a candidate enhancer is by placing its sequence in a plasmid with a minimal promoter and a reporter gene and delivering the plasmid to cell lines or embryos. The reporter gene's activity is then assessed by, for example, microscopy (for fluorescent proteins), *in situ* hybridisation or by measuring enzymatic activity (e.g. luciferase assays). However, these approaches do not allow for genome-wide screening, due to the time required to test each sequence (reviewed in Shlyueva et al., 2014).

Several groups have used a barcode-based technique which allows for the testing of multiple sequences simultaneously, in which candidate enhancers are cloned upstream of a minimal promoter and a reporter gene containing a unique barcode. RNA-seq is performed on cells with the plasmid and, with the different barcode for each candidate sequence, quantitative analysis can be performed to determine the level of expression driven by each candidate enhancer (Nam and Davidson, 2012, Patwardhan et al., 2012). However, this technique only allows the testing of hundreds or a few thousand sequences, due to the need to synthesise and clone all the different barcodes.

Arnold and colleagues developed a genome-wide enhancer screen, by placing the candidate sequences downstream of the minimal promoter. With this configuration, a functional enhancer will enhance its own expression, and it can be quantitatively measured by RNA-seq. This approach does not require different barcodes and has a simplified cloning step, allowing for the testing of millions of sequences (Arnold et al., 2013). The setback of approaches like these is that they can only be done in cell types which can be transduced, limiting its use in living

organisms, where developmental processes may alter the functionality of an enhancer (reviewed in Shlyueva et al., 2014).

To assess the possible enhancer activity of predicted sequences in Xenopus or zebrafish embryos, several groups have used Tol2-mediated transgenesis (Kawakami et al., 2004, Allende et al., 2006, Fisher et al., 2006, Hamlet et al., 2006, Kawakami, 2007, Bessa et al., 2009, Loots et al., 2013). In Tol2 transgenesis experiments, the predicted enhancer sequence is cloned upstream of a minimal promoter driving expression of, for example, GFP. The construct is then co-injected with Tol2 transposase mRNA, which will facilitate integration in the genome and, if the sequence is actually an enhancer, GFP expression should be observed in the same tissues and at the same stages as the gene that the enhancer endogenously regulates.

1.1.9 ChIP-seq

ChIP is a technique used to detect DNA-protein interactions *in vivo* and can be applied to, for example, determine protein and histone modification distribution across the genome, to then predict enhancer regions (as described extensively in the previous sections). It was first performed by Gilmour and Lis in *E.coli* to detect which genes were bound by RNA polymerase in *E.coli* (Gilmour and Lis, 1984). Cells were irradiated with UV light to covalently link DNA to bound RNA polymerase and an anti-RNA polymerase antibody was used to pull down DNA-protein complexes. This DNA could then be purified and used in hybridisation assays, to determine which sequences were bound by RNA polymerase. Currently, the most common crosslinking agent is formaldehyde, first used by Solomon and colleagues in *D. melanogaster* (Solomon et al., 1988).

ChIP-on-chip (or ChIP-chip) was first developed by Blat and Kleckner to study cohesin-bound locations in the yeast chromosome III (Blat and Kleckner, 1999), closely followed by three different groups which applied the technique to genome-wide microarrays (Ren et al., 2000, Iyer et al., 2001, Lieb et al., 2001).

With the advent of next-generation sequencing, ChIP-sequencing (ChIP-seq) was developed, allowing to determine the DNA sequence of the pulled down fragment with five groups publishing the method within four months of each other (Albert et al., 2007, Barski et al., 2007, Johnson et al., 2007, Mikkelsen et al., 2007, Robertson et al., 2007). Several methods have been developed based on ChIP-seq, such as ChIP-exo, which increases the experiment's resolution. This method, developed by Rhee and Pugh, uses an exonuclease to degrade the DNA up to the protein binding site, enabling the identification of the exact DNA sequence to which the protein of interest binds (Rhee and Pugh, 2012).

1.2 Xenopus tropicalis

In the rapidly developing embryo gene regulation is extremely dynamic, with diverse sets of genes being activated in different cells-types, so to allow differentiation into the multiple tissues which will give rise to the adult animal. This makes it an optimal system to study transcriptional regulation.

Xenopus tropicalis is being increasingly used in developmental biology studies, due to being diploid (compared with the allo-tetraploid *Xenopus laevis*), the ease of embryo handling and because most techniques used in *X. laevis* can be easily adapted for *X. tropicalis* use (Showell and Conlon, 2009). *Xenopus* has the advantage of producing large synchronously fertilized clutches with extrauterine development (Wheeler and Brandli, 2009) that allow microinjections and other physical manipulations to be performed at early cleavage stages, all year round. Having large numbers of synchronised cells allows the study of time-dependent processes, like transcriptional regulation, by next-generation techniques which require large quantities of starting material. The *X. tropicalis* genome is almost fully sequenced and research also benefits from the extensive expressed sequence tag (EST) libraries available (Showell and Conlon, 2009, Hellsten et al., 2010).

The first 12 divisions in *X. tropicalis* embryos are rapid and synchronous and all cellular processes are regulated by maternally deposited mRNAs and proteins. After the 12th division the mid-blastula transition (MBT) occurs and the embryonic control shifts to the zygotic genes with the first wave of zygotic transcription, as part of the maternal-to-zygotic transition (MZT) (Tadros and Lipshitz, 2009). At the MBT, as well as the activation of the zygotic genome, there is desynchronisation of the cell cycles, which become much longer, and cells become motile (Newport and Kirschner, 1982a). All these changes are believed to be triggered by the titration

relative to DNA of a maternally deposited component in the embryo (Newport and Kirschner, 1982b). Collart and colleagues reported four DNA replication factors, Drf1, Cut5, RecQ4 and Teslin, as being these hypothesised titrated proteins (Collart et al., 2013). The authors showed that these proteins are essential for MBT to occur and that their overexpression leads to a delayed MBT.

1.2.1 Dynamics of early transcription in *Xenopus tropicalis*

It is already known that zygotic transcription starts before the MBT (Yang et al., 2002, Skirkanich et al., 2011, Tan et al., 2013). Owens and colleagues measured the mRNA expression profiles from fertilisation until 66 hours post fertilisation (hpf) at 30-minute intervals (Collart et al., 2014, Owens et al., 2016). Importantly, in this study a method was developed to perform absolute normalisation on RNA-seq data, allowing the calculation of absolute transcript numbers in an embryo. This, in turn, allows the calculation of each gene's net transcription rate (Owens et al., 2016).

Both of these studies showed how dynamic transcription is in the developing embryo (Figure 4), which raises the question of what are the mechanisms that regulate this dynamic process.

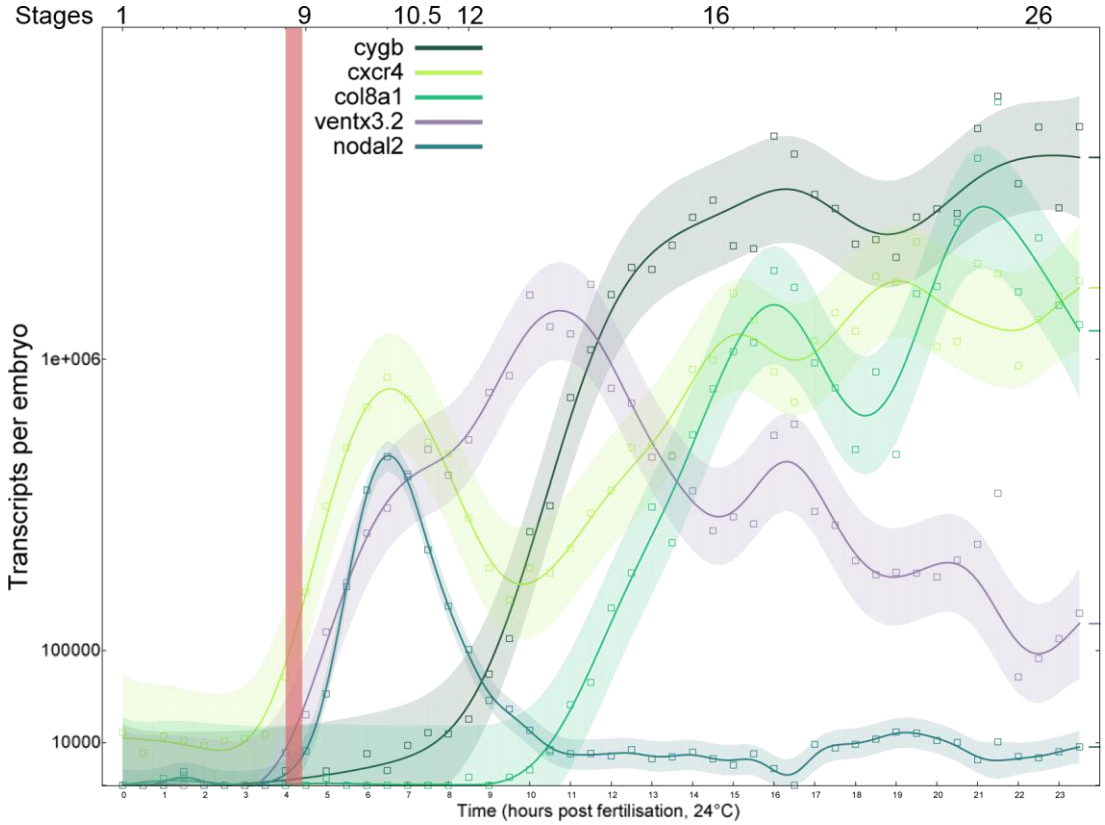


Figure 4 - Example of genes' transcription profiles in the first 23.5 hours of development.

(Data from Owens *et al.*, 2016). The y-axis represents the absolute number of transcripts per embryo and the x-axis the hours post fertilization. The pink bar marks the MBT; shaded area represents 95% confidence interval.

1.2.2 Enhancers and histone modifications in *Xenopus tropicalis* embryos

Hontelez and colleagues created epigenome maps at five developmental stages (Stages 9, 10.5, 12, 16 and 30) in *X. tropicalis*, for H3K27me3, H3K4me1/3, H3K9ac, H3K36me3, H3K9me2/3 and H4K20me3, plus binding maps for RNAPII and p300 and a DNA methylome map. p300 binding appears to vary between the assessed time points and different families of transcription factors associate with p300 at different developmental stages. They showed that most histone modifications are maternally defined and that p300 binding requires zygotic transcription – they blocked transcription by using α-amanitin and 85% of p300 regions disappeared (Hontelez et al., 2015). Even though they sampled five separate time points, the time intervals are too long to calculate dynamics. For example, for very dynamic genes, such as cxcr4 or ventx3.2 (Figure 4), if only those time points had been sampled we would not have a full correct picture of gene's transcription dynamics.

1.2.3 p300 in *Xenopus* embryos

p300 protein is present in *X. laevis* eggs (Wuhr et al., 2014), but its levels have not been determined in *X. tropicalis*. We do know its mRNA is maternally deposited. Figure 5 shows the expression profile for ep300, the gene that encodes p300, in absolute transcript numbers per embryo (Data from Owens *et al*., 2016). ep300 mRNA is present in the oocyte, its levels start increasing at around 4 hpf and it has a second wave of activation at around 13 hpf.

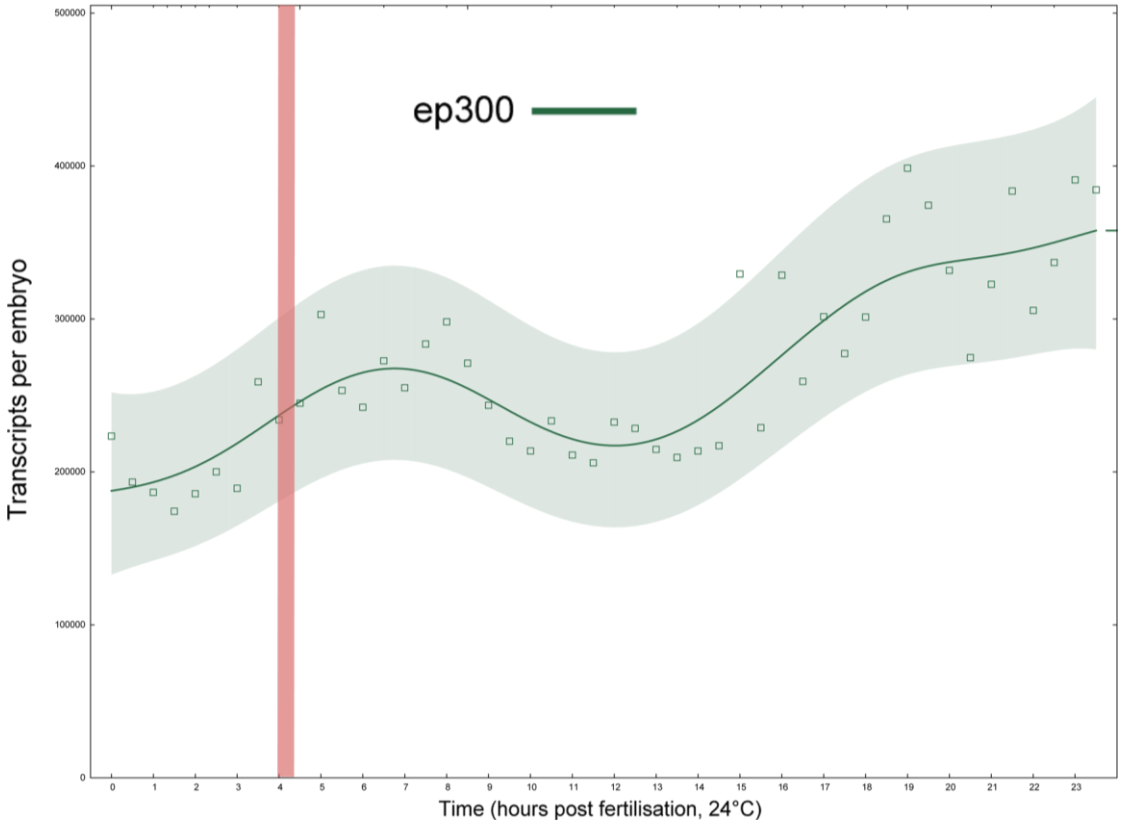


Figure 5 - ep300 RNA abundance in the first 23.5 hours of development.

(Data from Owens *et al*., 2016). The y-axis represents the absolute number of transcripts per embryo and the x-axis the hours post fertilization. The pink bar marks the MBT; shaded area represents 95% confidence interval.

1.3 Thesis Aim

The main aim of this project is to understand the dynamics of enhancer usage and how they relate to the dynamics of gene transcription, using the early developmental stages of *Xenopus tropicalis* as a model and p300 as an indicator molecule.

Data from the lab (Collart et al., 2014, Owens et al., 2016) showed how dynamic early transcription is, with genes being activated and deactivated in short time windows. This suggests that the mechanisms actually regulating transcription should display similar dynamic behaviour. As transcription is proposed to be, in part, regulated via enhancers, I hypothesised that these dynamics should be detectable in data associated with the enhancer regions. p300 has been widely used to predict enhancers (Wang et al., 2005, Heintzman et al., 2007, Heintzman et al., 2009, Visel et al., 2009a, Wang et al., 2009, Blow et al., 2010, May et al., 2011, Rada-Iglesias et al., 2011, Attanasio et al., 2013, Hontelez et al., 2015) and there were indications from p300 ChIP-seq sparse time series data from the Veenstra lab (Hontelez et al., 2015) that p300 binding is dynamic in early development. However, this data is too coarse grained to match up effectively with the observed transcriptional dynamics. Consequently, this project focused on generating high resolution temporal data to gain knowledge of the dynamic patterns of enhancer usage, which is essential to understanding how gene regulation is achieved.

Chapter 2. Materials & Methods

2.1 Embryo *in vitro* fertilisation

Animals were housed and all procedures performed according to the Animals (Scientific Procedures) Act 1986 (UK).

2.1.1 Oocyte collection

Adult *X. tropicalis* females were primed 24 hours before use with a subcutaneous injection into the dorsal lymph sac of 10 units of human chorionic gonadotropin (hCG, Sigma-Aldrich) in 100µL sterile water. Three to four hours before use, females were boosted in order to induce ovulation, with an injection of 200 units of the same hormone in 100µL sterile water. Oocytes were then collected from the female by gently pressuring the lower area of the abdomen, into a petri dish with one drop of 1x Marc's modified Ringers (MMR).

2.1.2 Sperm collection

Two males were killed by immersion in an overdose of Ethyl 3-aminobenzoate methanesulfonate (0.2%, MS222, Sigma-Aldrich) for at least 15 min at room temperature, followed by decapitation and destruction of the brain and spinal cord by double-pithing, as required by the schedule 1 guidelines. Testes were isolated through an incision in the abdominal wall and macerated in Leibovitz's L-15 medium (Sigma-Aldrich) with 10% fetal bovine serum (FBS, Sigma-Aldrich).

2.1.3 Fertilisation and embryo development

The testes suspension was used to fertilize the eggs from one female. After 4 to 5 min the eggs were flooded with 0.05x MMR. 10 min later, embryos were de-jellied with 2% cysteine hydrochloride (Sigma-Aldrich) in 0.05x MMR, pH8, for approximately 5 min and then washed in 0.05x MMR. Embryos were cultured at 26^{o}C in 0.05% MMR with gentamycin (100µL/mL, Sigma-Aldrich) until appropriate Niewkoop and Faber stage. For time series experiments, division times were recorded for at least the first 4 divisions, as well as approximate stage at time of collection.

2.2 Chromatin Immunoprecipitation (ChIP)

For ChIP the protocol by Gentsch and colleagues was followed, with some alterations (Gentsch and Smith, 2014, Gentsch et al., 2015). Recipes for solutions used are presented at 2.5 – Solutions.

2.2.1 Chromatin cross-linking

An appropriate amount of embryos (depending on the developmental stage) was collected, washed twice in 0.05x MMR and fixed in 1% formaldehyde (freshly opened capsule, Sigma-Aldrich) in 0.05x MMR for 25 minutes at room temperature. Fixed embryos were washed three times with ice cold 0.05x MMR and batches of less than 250 embryos were transferred to 2mL tubes and snap-frozen in liquid nitrogen. Embryos were stored at -80^{o}C for future use. For time series experiments with 30-minute time intervals, this procedure was performed with the assistance of Brook Cooper and Elena De Domenico due to time overlaps between performing the washes and the next time point embryo collection.

2.2.2 Chromatin extraction

Nuclei isolation and chromatin extraction was performed in batches of no more than 50-80 embryos. Fixed embryos were homogenised in E1 and centrifuged at 1000g for 2 minutes at 4^{o}C. The supernatant and lipids attached to the wall were discarded. The pellet was resuspended in E1 and incubated on ice for 10 min, followed by centrifugation and discarding of supernatant as before. The previous two steps were repeated twice with E2 and once with E3.

2.2.3 Chromatin fragmentation

At this point the pellet is formed of cross-linked nuclei, which were resuspended in 150µL of E3. Chromatin was fragmented by sonication in a Bioruptor® Plus (Diagenode) for 30 cycles, with 30/30 sec On/Off times. Fragmented chromatin was transferred into pre-chilled tubes, centrifuged at 15,000g for 5 min at 4ºC and the clear supernatant was transferred to a clean pre-chilled tube. 3-5% of the shredded chromatin was collected to use as input (non-immunoprecipitated chromatin) and stored at 4ºC.

2.2.4 Chromatin immunoprecipitation

13 µg p300 antibody (Santa Cruz Biotechnology, catalog number: sc-585) were added to the chromatin and the mixture was incubated overnight on a rotator (10 rpm) at 4ºC. 120 µL Dynabeads® Protein G (Invitrogen) per sample were washed for 5 min in E3, at 4ºC. The washed beads were added to the chromatin and antibody mixture and incubated for 4 hours on a rotator (10 rpm) at 4ºC.

From this point onward, a magnetic rack was used to assemble the beads/antibody/chromatin complexes at the bottom of the tubes before discarding the supernatant. Samples were washed 10 times with pre-chilled RIPA buffer, with 5 min incubations on a rotator (10 rpm) at 4ºC in between washes. Samples were washed once more, this time in pre-chilled TEN buffer for 5 min on a rotator (10 rpm) at 4ºC. Pellets were resuspended in 50µL of TEN buffer and the suspension was transferred to a new tube. Samples were centrifuged at 1000g for 1min at 4ºC and the supernatant was discarded. 100µL SDS elution buffer was added to the beads and the mixture was shaken for 15 min in a thermomixer set to

1000 rpm and 65°C. Samples were centrifuged at 15,000g for 30 sec at room temperature. Supernatants were transferred to a new tube. Last three steps were repeated. ChIP samples had 200 µL by the end of these steps.

2.2.5 Reverse cross-linking

SDS elution buffer were added to input samples to 200 µL. ChIP and input samples were supplemented with 20 µL 5M NaCl and were incubated overnight at 65°C. 200 µL TE buffer and 8 µL RNAse A (final concentration 200 µg/mL, Invitrogen) were added and incubated for 1 hr at 37°C. 4 µL proteinase K (final concentration 200 µg/mL, Ambion) were added and incubated for 2 to 4 hours at 55°C.

2.2.6 DNA purification and sequencing

DNA was purified using the MinElute PCR Purification Kit (Qiagen) according to the manufacturer's protocol and eluted in ~20 µL.

The library preparation step and sequencing was performed by GATC Biotech or the High-Throughput Sequencing team at The Francis Crick Institute/Mill Hill Lab.

2.3 Pilot p300 ChIP-seq time series collection

X. tropicalis embryos were generated as described in section 2.1. Fertilized embryos were allowed to develop for 7 hours and were then collected every 30 minutes, until 10.5 hours post fertilization, with the help from Brook Cooper. Embryos were processed as described in section 2.2.

2.4 Long p300 ChIP-seq time series collection

X. tropicalis embryos were generated as described in section 2.1. Fertilized embryos were allowed to develop for 5 hours and were then collected every 30 minutes, until 17.5 hours post fertilization, with the help of Elena De Domenico. Embryos were processed as described in section 2.2.

2.5 Solutions

Most solutions were prepared by The Francis Crick Institute/Mill Hill Lab services, including 50 mM HEPES (pH 7.5), 1 mM EDTA, 50 mM Tris-HCl (pH 8.0).

10x Marc's Modified Ringers MMR

1 M NaCl

20 mM KCl

20 mM CaCl2

10 mM MgSO4

50 mM HEPES (pH 7.5)

pH adjusted to 7.5 and solution sterilised by autoclaving.

E1

50 mM HEPES (pH 7.5)

150 mM NaCl

1 mM EDTA

10% glycerol

0.5% Igepal CA-630

0.25% Triton X-100

0.2mM PMSF in 100% EtOH

1mM DTT

1 protease inhibitor tablet (Roche) per 10mL

(PMSF, DTT and protease inhibitor should only be added right before use)

E2

10 mM Tris-HCl (pH 8.0)

150 mM NaCl

1 mM EDTA

0.5 mM EGTA

0.2mM PMSF in 100% EtOH

1mM DTT

1 protease inhibitor tablet (Roche) per 10mL)

(PMSF, DTT and protease inhibitor should only be added right before use)

E3

10 mM Tris-HCl (pH 8.0)

150 mM NaCl

1 mM EDTA

1% Igepal CA-630

0.25% Na-Deoxycholate

0.1% SDS

0.2mM PMSF in 100% EtOH

1mM DTT

1 protease inhibitor tablet (Roche) per 10mL)

(PMSF, DTT and protease inhibitor should only be added right before use)

RIPA buffer

50 mM HEPES (pH 7.5)

500 mM LiCl

1 mM EDTA (pH 8.0)

1% Igepal CA-630

0.7% Na-deoxycholate

SDS elution buffer

50 mM Tris-HCl (pH 8.0)

1 mM EDTA

1% SDS

TEN buffer

10 mM Tris-HCl (pH 8.0)

1 mM EDTA (pH 8.0)

150 mM NaCl

All solutions were prepared in double-distilled water

2.6 ChIP-seq analysis

Unless otherwise stated, data analysis was done using command line and Python 3.5 (www.python.org) and all graphs were produced with python packages matplotlib (Hunter, 2007) and seaborn (seaborn.pydata.org).

2.6.1 Sequencing reads analysis

Raw sequencing files were first analysed using FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to determine the quality of the sequencing reads. They were then mapped to the *X. tropicalis* genome assembly 7.1 (Hellsten et al., 2010, Karpinka et al., 2015) using bowtie2 (Langmead and Salzberg, 2012) using default parameters (used **-X 700** to align paired-end reads).

For paired-end sequencing, reads not mapped to proper pair were removed (**samtools view -f 2**) (Li et al., 2009). Non-unique reads were filtered out by using **grep -v XS:i**. Bedtools (Quinlan and Hall, 2010) bamtobed was used to convert the BAM files to BED files (or BEDPE for paired-end sequencing) and genomecov was used to generate the BEDGRAPH files and normalise reads by total number of mapped reads. wigToBigWig was used to convert the BEDGRAPH file to bigwig (https://genome.ucsc.edu/goldenpath/help/bigWig.html), to be displayed in a genome browser.

2.6.2 Peak calling

MACS2 (Model based analysis of ChIP-seq 2) (Zhang et al., 2008) was used to perform peak calling, to identify regions with enriched p300 binding over background, with an FDR of 0.1% and the following parameters: **--SPMR -q 0.001 --call-summits** (**-f BAM** for single-end and **-f BAMPE** for paired-end data).

2.6.3 p300 region-Gene assignment

p300 regions were assigned to their closest gene using **annotatePeaks.pl** from HomerTools (Heinz et al., 2010)

2.6.4 p300 region clustering

p300 regions were clustered using the python package scipy (https://www.scipy.org) clustering. Data was kmeans clustered, with the k value being heuristically determined, by testing several values and determining which one led to the best region separation, without overclustering. The average time point at which the cluster reached its maximum expression was calculated and clusters were ordered by that value. Clusters with identical maxima were manually ordered.

2.6.5 Random region generation

Random regions were generated using **bedtools random**, with the appropriate number of required regions. Random regions were then annotated as mentioned on section 2.7.3.

2.6.6 Genome browser and heatmaps

Genome browser profiles were generated with fluff (Georgiou and van Heeringen, 2016). Read density heatmaps were generated using deepTools (Ramirez et al., 2016), **computeMatrix reference-point** with parameters **--referencePoint center -a 5000 -b 5000** in order to centre heatmaps around the middle of the p300 region and to count reads in a 10 kb area around it. Heatmap was then plotted using **plotHeatmap --sortRegions no**, so regions maintained the clustering ordering.

Chapter 3. Identification of Candidate Enhancers and p300 Dynamics

3.1 Introduction

Creating a dataset of candidate enhancer regions is an essential tool in the genetic and epigenetic study of any model organism. Having such a tool allows further study on how specific enhancer mutations may lead to a disorder and also opens the possibility to modulate gene expression by modifying enhancers of interest. p300 has been used extensively to predict active enhancer regions in different model organisms (Wang et al., 2005, Heintzman et al., 2007, Heintzman et al., 2009, Visel et al., 2009a, Wang et al., 2009, Blow et al., 2010, May et al., 2011, Rada-Iglesias et al., 2011, Attanasio et al., 2013), including *Xenopus tropicalis* (Hontelez et al., 2015).

To understand transcriptional regulation – which underpins much of development – it is important to understand the temporal dynamics of enhancer behaviour. p300 has been assayed at different developmental stages; however, to date, no experiment has been performed to determine how p300 binding changes during short time intervals. Currently, the most frequently sampled p300 datasets have been assayed at developmental stages several hours apart. Having high resolution time data would allow the determination of the fast dynamics that may be involved in the fine-tuning of gene expression in the early embryo.

Work presented in this chapter aimed to create a dataset of candidate enhancer regions in early *Xenopus tropicalis* development, determine their usage dynamics, using p300 as a proxy for this, and compare the identified regions with previously published histone modifications and RNAPII data. In order to achieve

this, a pilot p300 ChIP-seq time series was performed in *Xenopus tropicalis* embryos, from 7 to 10.5 hpf, at 30 min time intervals, followed by a longer experiment, from 5 to 17.5 hpf, also at 30 min time intervals (Figure 6).

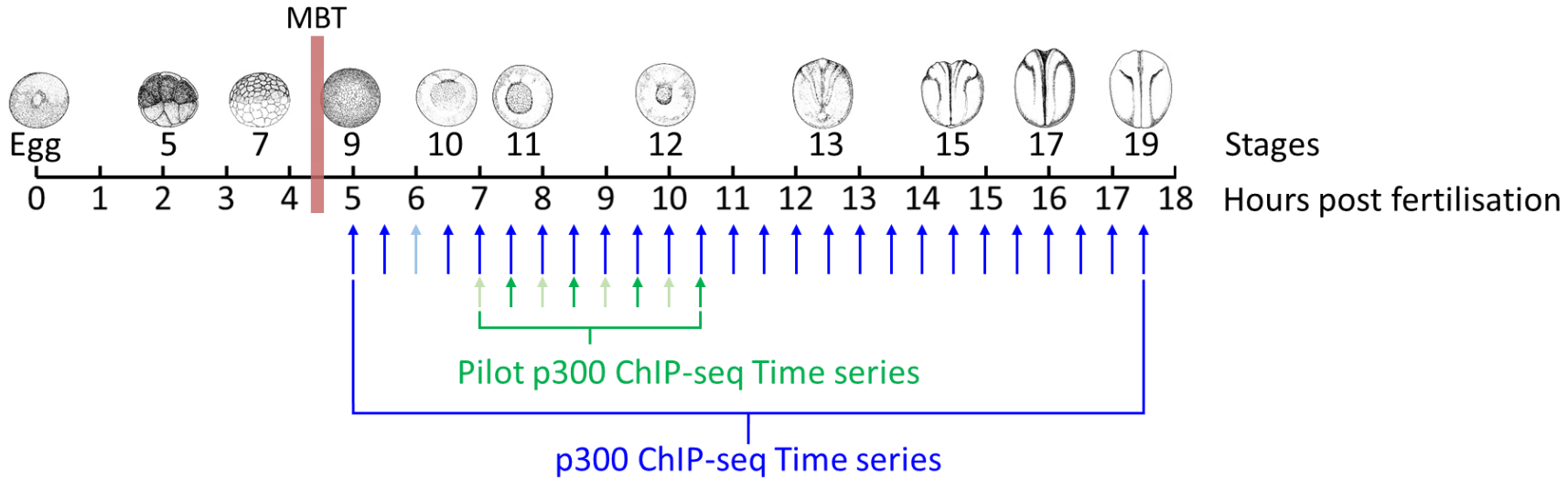


Figure 6 – Time series collection diagram.

Xenopus tropicalis embryos from one synchronous clutch were collected for a pilot (green) and a long (blue) p300 ChIP-seq time series experiments, at the time points marked. Ligther coloured arrows mark time points not sequenced. Pink bar marks MBT.

3.2 p300 ChIP-seq

3.2.1 p300 ChIP-seq pilot time series

3.2.1.1 Embryo Collection

Xenopus tropicalis embryos were collected with the help from Brook Cooper, for a pilot p300 ChIP-seq time series. Embryos were collected as described in Materials and Methods, at the times shown on Figure 6 (green), from 7 to 10.5 hpf, with the number of embryos per sample described on Table 1. The number of embryos needed for the ChIP-seq decreases with the increasing cell number per embryo. This clutch had a division time during the cleavage stages of 17 minutes. Samples will be referred to by their time (e.g. Sample collected at 7.5 hpf will be referred to as Sample 7.5).

Hours post fertilisation (hpf)	Number of embryos
7	75
7.5	74
8	75
8.5	57
9	55
9.5	53
10	53
10.5	55

Table 1 – Time and number of *X. tropicalis* embryos collected from a single synchronous clutch, for a pilot p300 ChIP-seq time series.

3.2.1.2 Sequencing results

p300 ChIP was performed on all samples as described in Materials and Methods. All samples were sent to GATC Biotech for library preparation and paired-end (51 bp) sequencing in an Illumina HiSeq. It was requested that they first sequenced samples 7.5, 8.5, 9.5 and 10.5 hpf and the matched inputs, to avoid sequencing all samples if the quality of the experiment was not sufficient. After receiving positive results from that set of samples, it was requested that they sequenced the rest of the samples and inputs, however GATC Biotech reported that they were not able to perform library preparation on samples 7, 8, 9 and 10 hpf. Therefore, the pilot time series was analysed with a 1 hour time interval.

ChIP-seq analysis was performed as described in Materials and Methods. FastQC analysis was performed to determine the quality of the sequencing reads and all samples had very high quality. After read alignment, about 60% of read pairs in each sample mapped uniquely. Table 2 summarises sequencing results.

Samples	Number of read pairs (Millions)	Uniquely mapped pairs (Millions)
7.5	41.9	25.6 (61.0%)
8.5	22.5	13.7 (61.0%)
9.5	26	15.9 (61.2%)
10.5	30.7	18.8 (61.5%)
Input 7.5	47.5	28.6 (60.2%)
Input 8.5	46.3	29.6 (63.8%)
Input 9.5	44.5	27.7 (62.3%)
Input 10.5	49.1	27.6 (56.2%)

Table 2 – Number of sequencing and uniquely mapped read pairs in pilot p300 ChIP-seq time series.

Figure 7 represents the ChIP-seq data normalised by total number of uniquely aligned reads, per sample, in two loci with different p300 binding dynamics. The *mix1*/*mixer* locus has p300 binding at early time points, with that binding disappearing towards later development. The *lrrk1* locus shows one p300 region with maximum binding at 8.5 hpf and regions with later p300 binding.

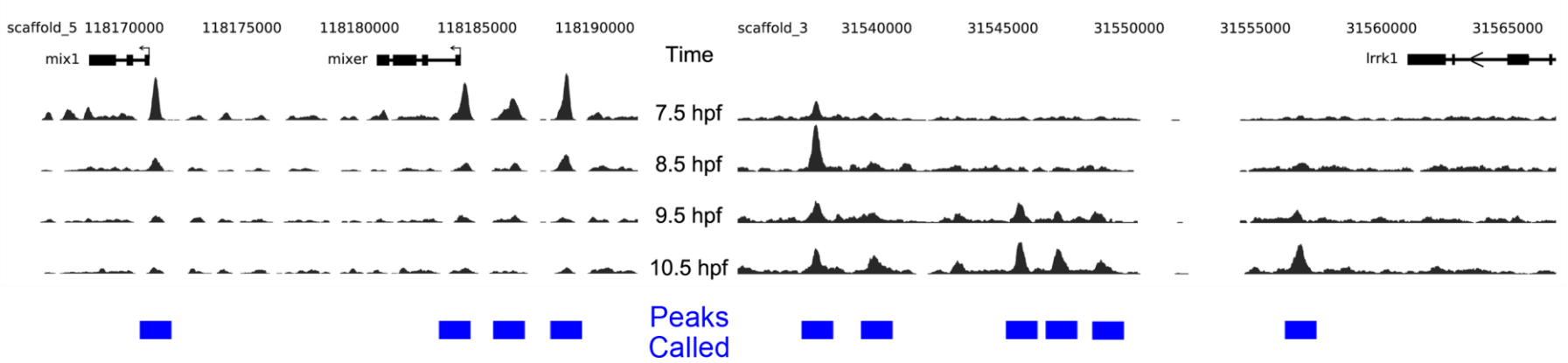


Figure 7 – Pilot p300 ChIP-seq time series.

Genome browser view of *mix1*/*mixer* and *lrrk1* loci and the p300 peaks called by MACS2 (blue boxes).

3.2.1.3 Peak calling

Peak calling was then performed using MACS2, as described in Materials and Methods, with a false discovery rate of 0.1%. An example of the peak regions called in one or more samples can be seen on Figure 7.

The majority of ChIP-seq experiments are done with single-end sequencing; in order to determine if paired-end sequencing has advantages, peak calling was performed both on the paired-end data and on the corresponding single-end data.

Samples	Peaks in Paired-end	Peaks in Single-end	Overlap paired vs single-end	Overlap top paired vs single-end
7.5h	12516	8344	65.9%	91.6%
8.5h	10657	6918	64.8%	92.2%
9.5h	8044	4738	58.0%	85.3%
10.5h	11570	6716	58.9%	85.8%

Table 3 – Pilot p300 ChIP-seq time series peak calling results.
Numbers of peaks called per sample and how the ones in the single-end data compare with the ones called in the paired-end data. Overlap represents the percentage of paired-end peaks also present in the single-end data.

Table 3 shows the number of peaks called for each sample in the paired-end and in the single-end data and the percentage of peaks called in the paired-end that are also called in the single-end data (overlap paired vs single-end). About 40% of peaks are lost in the single-end data, most likely due to the actual fragment size in single-end data being unknown, leading to a worse performance from MACS2. To determine how the highest scored peaks behaved, peaks were sorted by their MACS2 score and the top half were selected. These were then compared to the peaks called in the single-end data, with the results described on Table 3. About 90% of the highest scored peaks are maintained when analysing

only single-end data, therefore, the majority of peaks missed in the single-end data have a low score and are more likely to be false positives.

With the goal of performing a longer p300 ChIP-seq time series, it was important to determine if input samples for every time point needed to be sequenced in order to have reliable peak calling. To address this, peaks in sample 7.5 were called using input 10.5, instead of input 7.5 (the input initially used); 92.5% of the peaks were maintained and, looking only at the top half of the peaks, this number increased to 98.7%. This indicates that it is not essential to have input samples for all time points because the vast majority of peaks are maintained when an input from embryos 3 hours later is used. Therefore, I decided not to sequence all input samples in subsequent experiments.

3.2.1.4 Pilot time series correlation

Figure 8 represents the Pearson correlation coefficient between pairs of samples in the pilot time series. As expected, samples are more closely correlated with the adjacent sample than to samples further away in time, which shows the high quality of the data and also that time dynamics are present in the data – because neighbouring time points are better correlated, if there were no time dynamics this effect would not be present.

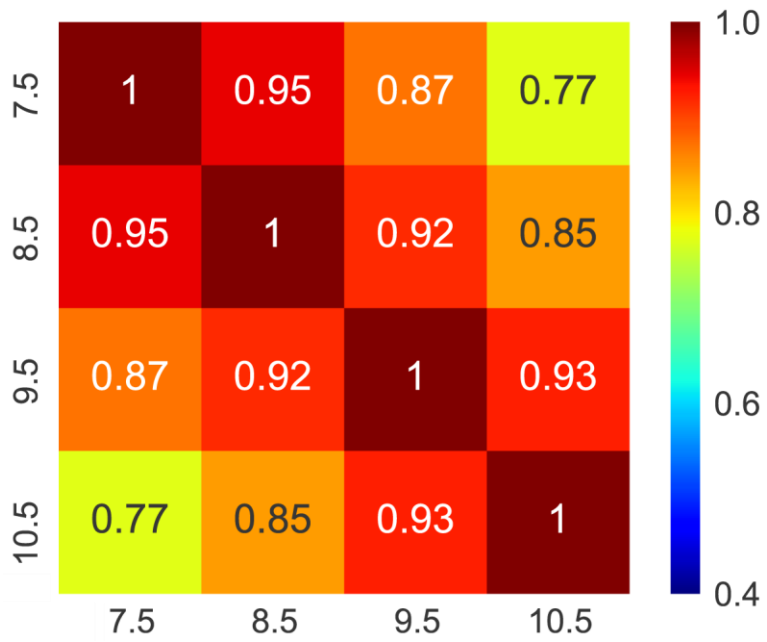


Figure 8 – Pilot p300 ChIP-seq time series correlation.
Pearson correlation between pairs of pilot p300 ChIP-seq time series samples.

3.2.1.5 Conclusions from pilot p300 ChIP-seq time series

The pilot p300 ChIP-seq time series showed how, even with only 4 time points over 3 hours, it is possible to see dynamic p300 binding (Figure 7 and Figure 8). Paired-end sequencing yields higher peak numbers, however most of the highly scored peaks are maintained in single-end sequencing (Table 3). Given that paired-end sequencing costs nearly twice as much as single-end, subsequent ChIP-seq experiments were performed with single-end sequencing. Peak calling does not seem to be affected by using input from samples three hours apart, therefore, in subsquent ChIP-seq experiments, not all input samples were sequenced.

3.2.2 p300 ChIP-seq long time series

3.2.2.1 Embryo Collection

A much longer time series was collected, as shown on Figure 6 (blue), from 5 to 17.5 hpf, at 30 min time intervals, with the help from Elena De Domenico. Embryos were collected as described in Materials and Methods, with the number of embryos per sample described on Table 4. Sample 6 was lost during the ChIP experiment. As previously mentioned, earlier time points require a higher number of embryos due to the lower cell number per embryo. Approximate developmental stage was recorded during collection. This clutch had a division time of 18 minutes during the cleavage stages.

Hours post fertilisation	Number of embryos	Approximate Stage
5	100	9
5.5	100	9
6	90	9
6.5	90	10
7	75	10
7.5	74	11
8	70	11
8.5	60	11
9	55	11
9.5	55	11
10	55	12
10.5	55	12
11	50	12
11.5	50	12
12	50	12
12.5	50	13
13	40	13
13.5	40	14
14	40	14
14.5	40	15
15	40	16
15.5	40	16
16	40	17
16.5	35	17
17	35	18
17.5	35	19

Table 4 – Embryo collection for p300 ChIP-seq long time series.

Time, number of embryos and approximate developmental stage of *X. tropicalis* embryos collected for the p300 ChIP-seq long time series.

3.2.2.2 Sequencing results

p300 ChIP was performed on all samples as described in Materials and Methods. The High-Throughput Sequencing team at The Francis Crick Institute/Mill Hill Lab performed library preparation and single-end (51bp) sequencing in an Illumina HiSeq. Input samples were produced for all time points, however, due to cost considerations and given that it was shown in the pilot project that most peaks (~90%) are maintained when using inputs for other samples, only inputs 5, 7, 9, 11, 13, 15 and 17 were sequenced.

FastQC analysis showed that all samples had high-quality sequencing reads. Table 5 summarises sequencing results, showing about 50% of reads mapped uniquely in each sample.

Peak calling was then performed as described in Materials and Methods, with a false discovery rate of 0.1%.

Samples	Number of reads (Millions)	Uniquely mapped reads (Millions)	Number of peaks	Input sample used
5	50.9	22.9 (44.9%)	7152	5
5.5	61	30 (49.1%)	1256	5
6.5	67.5	33.9 (50.2%)	53	7
7	65.9	32.5 (49.3%)	297	7
7.5	66.9	33.7 (50.3%)	186	7
8	54.7	28 (51.3%)	68	9
8.5	46.2	22.9 (49.7%)	2086	9
9	54.3	27.6 (50.8%)	2721	9
9.5	53.1	27.1 (51.0%)	3980	9
10	61	31.8 (52.1%)	2173	11
10.5	41.9	19.5 (46.5%)	8238	11
11	37.5	18.7 (49.7%)	3349	11
11.5	46.9	23.9 (50.9%)	5965	11
12	51.7	26.4 (51.1%)	2167	13
12.5	62.5	32.7 (52.3%)	2479	13
13	50.2	26 (51.7%)	2166	13
13.5	35.2	19.2 (54.8%)	1384	13
14	46.2	25.1 (54.4%)	3181	15
14.5	57	30.8 (52.9%)	1364	15
15	37.8	20.2 (53.5%)	2786	15
15.5	42	23.6 (56.2%)	2049	15
16	43.5	24.3 (56%)	3633	17
16.5	45.8	25.7 (56.3%)	809	17
17	35.9	14.9 (41.6%)	2784	17
17.5	51.7	23.4 (45.3%)	242	17
Input 5	82.1	43.5 (53%)		
Input 7	67.9	37 (54.6%)		
Input 9	44.9	24.9 (55.5%)		
Input 11	41.2	22.5 (54.5%)		
Input 13	25.7	14.2 (55.4%)		
Input 15	37.7	20.4 (54.2%)		
Input 17	38.1	21.7 (56.7%)		

Table 5 – p300 ChIP-seq long time series sequencing.

Sequencing and uniquely mapped reads in the p300 ChIP-seq long time series, number of peaks called and which input sample was used for peak calling.

3.2.2.3 ChIP-seq normalisation

p300 ChIP-seq data normalised by the total number of uniquely mapped reads in each sample is represented on Figure 9A (using the *ventx* locus as an example). Samples 5.5 to 8 have generally lower peaks and less peaks called, compared to other samples (Table 5). This was visible across the whole genome (data not shown), showing it is not a region specific phenomenon. These samples appeared to have a higher percentage of background reads, i.e. reads which map apparently randomly across the genome.

Samples with higher background levels will have their peaks' read density underestimated because, all other things being equal, sequencing reads are spread all over the genome and not concentrated in binding regions, as they are in high quality samples. This would negatively impact quantitative analysis of the data. In order to overcome this issue, the data were normalised in a way that discounted background reads.

The following normalisation method was developed:

- Peak calling was performed as described above, giving a set of regions called in each sample.
- A peak file was created, with all the peaks from all samples. Any overlapping peaks or peaks with edges less than 49 bp away from a nearby peak were merged (with 51 bp reads, a read could be counted as aligning to two different peaks if those were less than 49 bp apart). Hereafter, this will be referred to as *the set of p300 regions*.
- For each sample, the number of reads aligning to the set of p300 regions was counted, irrespectively of whether a specific peak was

called in any given sample or not. That number was then used to derive the normalisation factor for its corresponding sample.

- In summary, each sample is normalised by the total number of reads that map to all peaks called in all samples (set of p300 regions).

In total, 17,414 p300 regions were identified.

Figure 9 shows a comparison of the data in the *ventx* locus: A representing the data normalised by total number of uniquely mapped reads, B representing the data normalised by total number of reads aligning to the set of p300 regions.

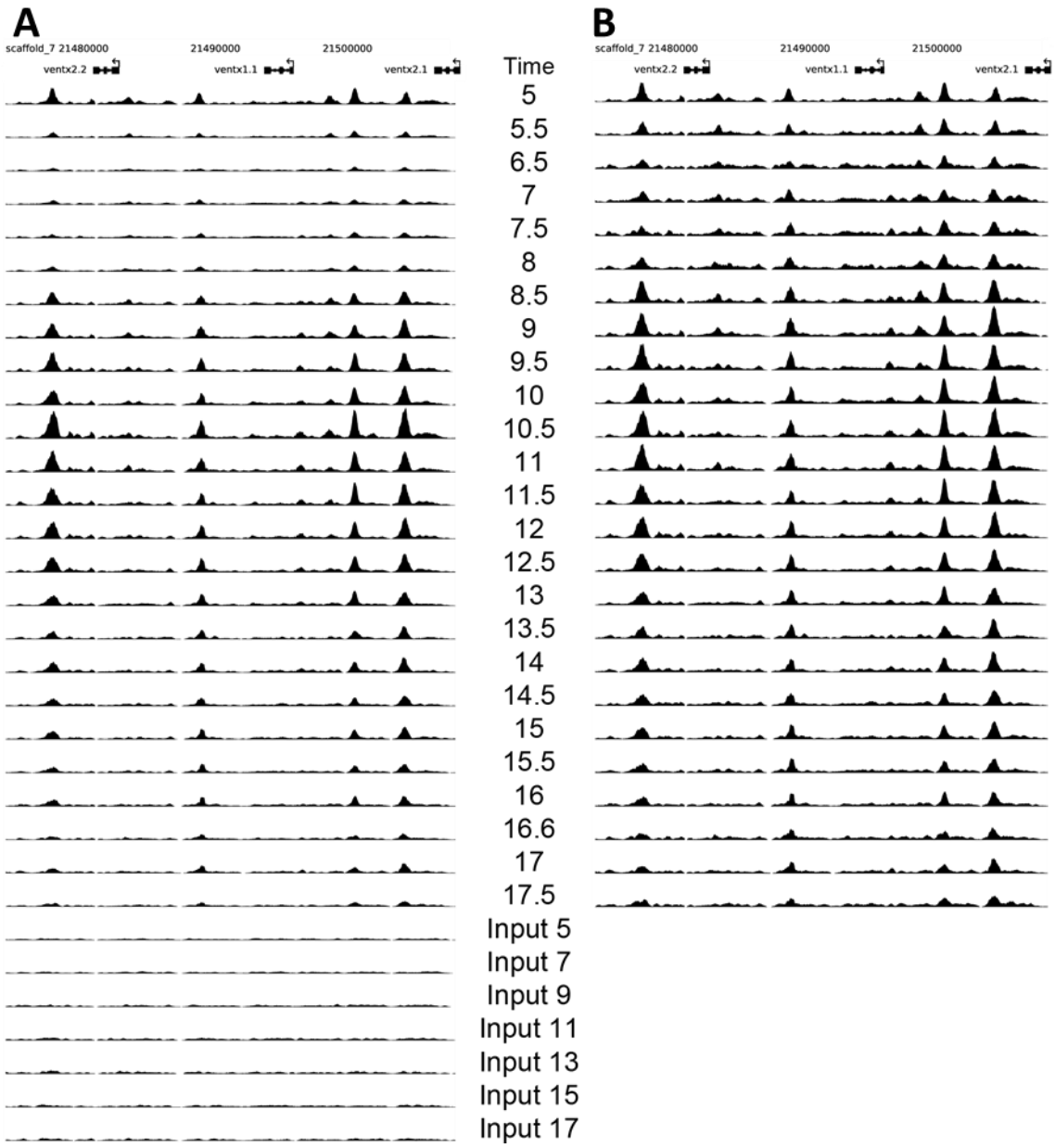


Figure 9 – p300 ChIP-seq long time series.

A – Genome browser view of the *ventx* locus, with p300 ChIP-seq reads normalised by the total number of uniquely mapped reads. B – Genome browser view of the same locus, with p300 ChIP-seq reads normalised by the total number of reads aligning to the set of p300 regions.

Having a set of p300 regions also allows the calculation of a sample's enrichment – which I define as the percentage of total mapped reads that align to the set of p300 regions. Table 6 summarises that data and samples 5.5 to 8, which seemed to have smaller peaks and higher background, do indeed have lower levels of enrichment. Figure 10 shows how the number of peaks called correlate with the sample's enrichment; the higher the latter, the more peaks MACS2 is able to call.

Samples	Number of peaks	Reads in set of p300 regions (Thousands)	Enrichment
5	7152	680	2.98%
5.5	1256	432	1.44%
6.5	53	390	1.15%
7	297	438	1.35%
7.5	186	436	1.30%
8	68	388	1.38%
8.5	2086	461	2.01%
9	2721	617	2.23%
9.5	3980	707	2.61%
10	2173	786	2.47%
10.5	8238	717	3.68%
11	3349	527	2.82%
11.5	5965	748	3.13%
12	2167	757	2.87%
12.5	2479	972	2.97%
13	2166	725	2.79%
13.5	1384	445	2.31%
14	3181	654	2.61%
14.5	1364	711	2.36%
15	2786	506	2.50%
15.5	2049	522	2.21%
16	3633	611	2.51%
16.5	809	462	1.79%
17	2784	465	3.12%
17.5	242	497	2.12%

Table 6 – p300 ChIP-seq long time series samples' enrichment.

Number of peaks called per sample, number of reads in the set of p300 regions and sample's enrichment.

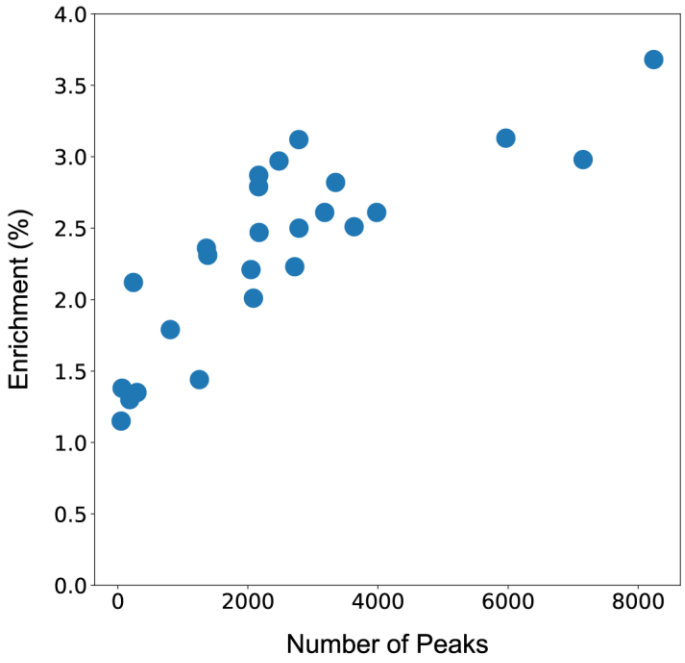


Figure 10 – Enrichment vs number of peaks called by MACS2 for p300 ChIP-seq long time series samples.

By normalising the samples by the read count in the set of p300 regions, the subsequent analysis is not dependent on varying background levels and varying quality of peak calling.

3.2.2.4 Pilot and long time series show high replicability

To determine how replicable p300 ChIP-seq is, the Pearson correlation coefficients between the pilot and long time series were calculated, at each time point in common. Clutch division times were slightly different (17 minutes in the pilot and 18 in the long time series), so all times had to be adjusted (Table 7). Samples correlated well between the two experiments, even though there is a slight time lag between the pairs (Table 7 and Figure 11).

Time in pilot time series	Adjusted time in long time series	Pearson correlation coefficient
7.5	7.94 ≈ 8	0.83
8.5	9	0.89
9.5	10.05 ≈ 10	0.9
10.5	11.11 ≈ 11	0.82

Table 7 – Pearson correlation coefficient between pilot and long p300 ChIP-seq time series.

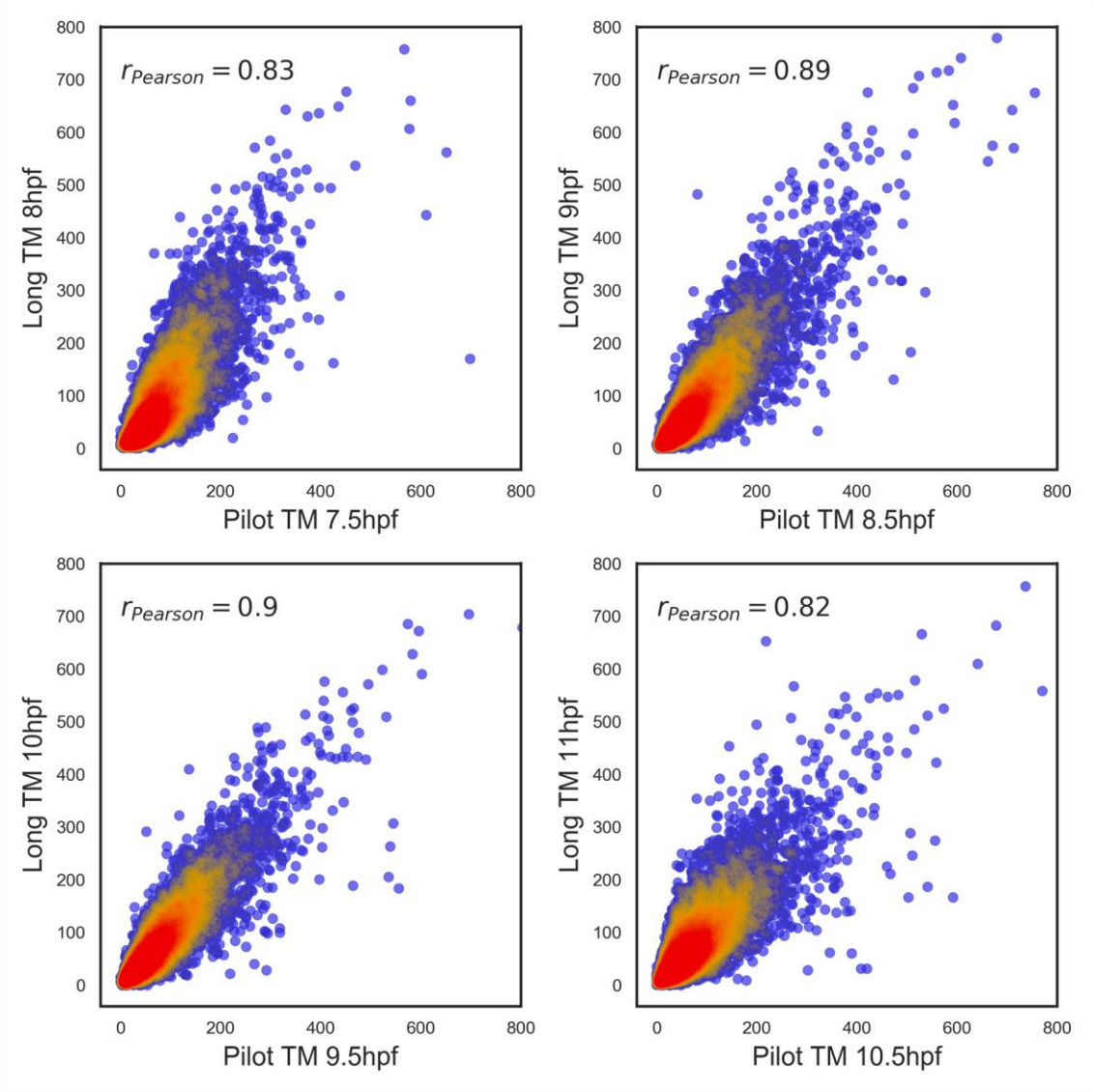


Figure 11 – Scatter plots of normalised p300 reads for pairs of pilot and long p300 ChIP-seq samples at equivalent times and their correlation.

Each dot represents one peak region; the y-axes represent the number of normalised reads in samples of the long time series and the x-axes in samples of the pilot time series. The colour represents the dot density. Times have been adjusted due to different clutch division times (Table 7). TM – Time series.

3.2.2.5 Time series adjacent time points correlate better than biological replicates

Pearson correlation coefficients were calculated between all pairs of samples in the long time series experiment (Figure 12). As expected, neighbouring samples are very highly correlated (>0.9), being even more correlated than biological replicates (from different clutches) (Table 7), which shows that the time interval is small enough for adjacent samples to function as biological replicates and to capture potentially important time dynamics.

Figure 12 also shows that time dynamics are present in the data, with the correlation coefficient steadily decreasing with increased time distance between the samples (e.g: 5 hpf vs 17.5 hpf correlation coefficient is 0.17.)

Sample 10.5 had a significantly lower correlation coefficient to adjacent time points so it was removed from further analysis.

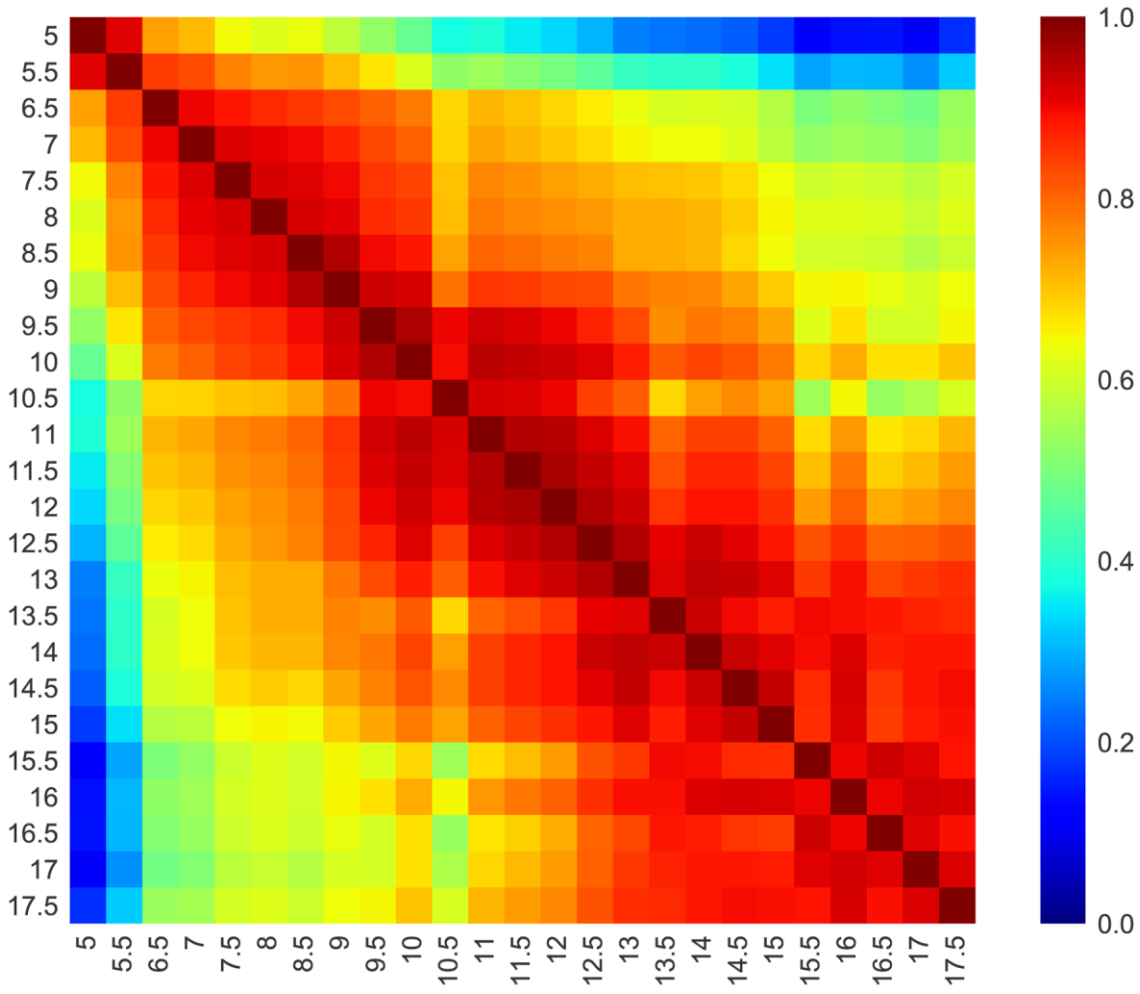


Figure 12 – p300 ChIP-seq long time series samples' correlation.

Pearson correlation between all pairs of p300 ChIP-seq long time series samples.

3.2.2.6 Gaussian processes and data filtering

Owens and colleagues (Owens et al., 2016) developed a method based on Gaussian processes to calculate a statistical model of time series data which, amongst other quantities, results in a median line of best fit for each RNA expression profile, and the corresponding 95% confidence intervals. This method estimates three parameters by maximising the marginal likelihood:

- σ_f^2, the signal variance, which measures the scale of the normalised read count for that region (how high are p300 levels at a given region).
- τ, the timescale or length-scale, which indicates how fast the signal can change. It is calculated based on the times the line of best fit crosses a given threshold.
- σ_n^2, the noise variance, which measures the noise around the calculated line of best fit and influences the size of the confidence intervals.

For further information on Gaussian processes and the equations used to calculate each parameter, I would direct the reader to section 3.3 (Gaussian Process Models of Gene Expression) of the Supplemental Information of Owens et al., 2016.

In order to perform dynamic quantitative analysis, this algorithm was adapted to the p300 ChIP-seq data generated in this project. The normalised read count in individual p300 regions was calculated genome-wide, for each time point. These values were then used to fit Gaussian processes and so calculate the smooth line that best describes the p300 binding profile for each p300 region. Figure 13 shows the genome browser view of an example p300 region and the

resulting curve that best describes the data, with the corresponding 95% confidence interval.

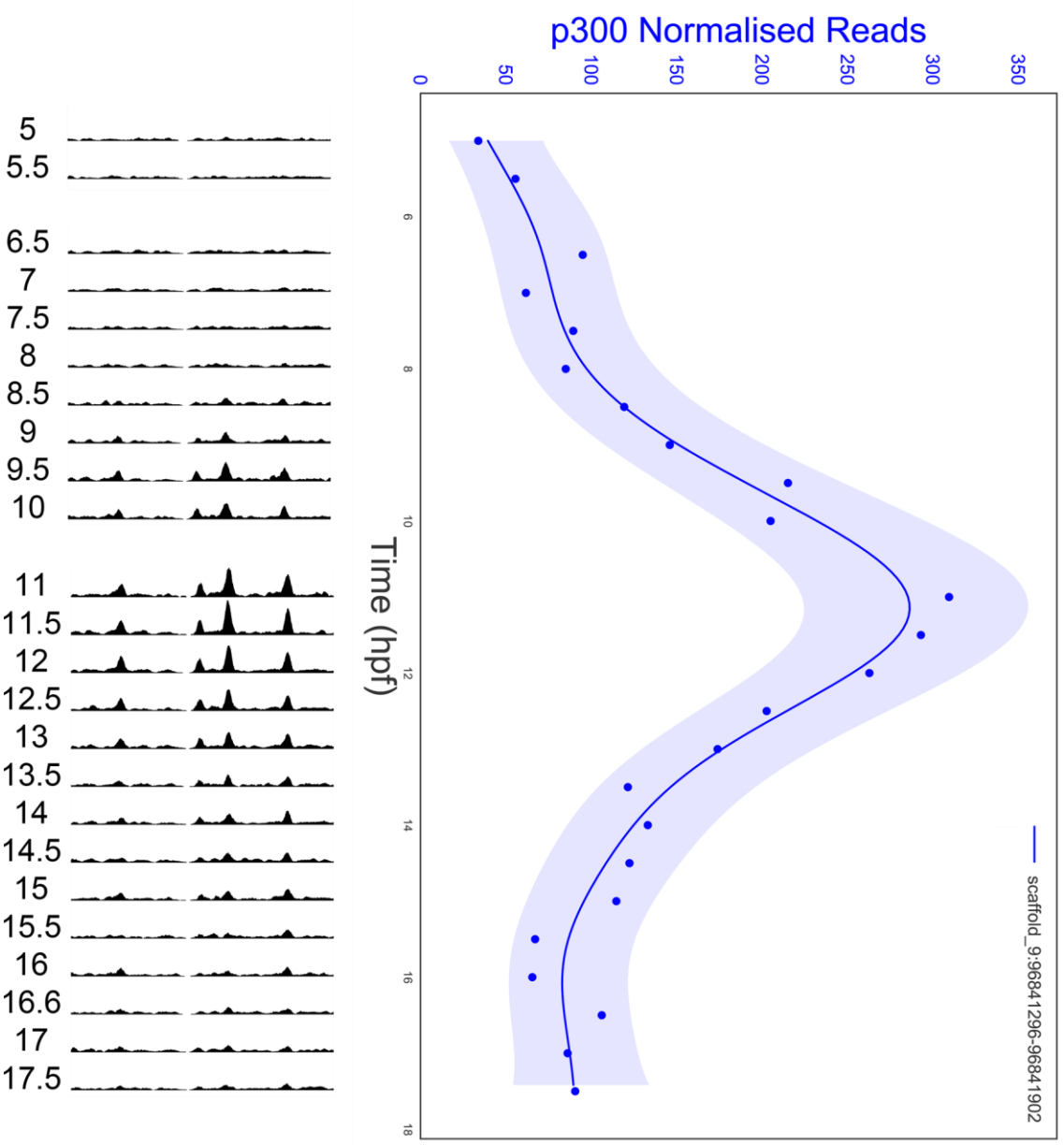


Figure 13 – Gaussian process method results.

Example genome browser view with one p300 region highlighted (blue box) and the resulting Gaussian process graph. Points represent the normalised read count for the highlighted region at each time point; curve marks the median line of best fit for that region and the shaded area is the 95% confidence interval.

Figure 14 shows some of the different profiles obtained, with some regions having maximum p300 binding at earlier stages and then decreasing during development; regions where binding increases during the time series; regions with more dynamic behaviours, with p300 binding increasing and decreasing during this time interval; and regions which are almost constant.

In order to remove regions with no obvious time signal, where reads vary significantly between adjacent samples, the signal-to-noise ratio (SNR) was calculated for each p300 region. The SNR was calculated using the Gaussian process parameters described above, with SNR = $\log(\sigma_f^2/\sigma_n^2)$. The filtering threshold was heuristically determined and was set at SNR > 3.8, which excluded 4842 regions (28%) (for example, the lower two profiles in Figure 14). All further analysis was done with the filtered set of p300 regions.

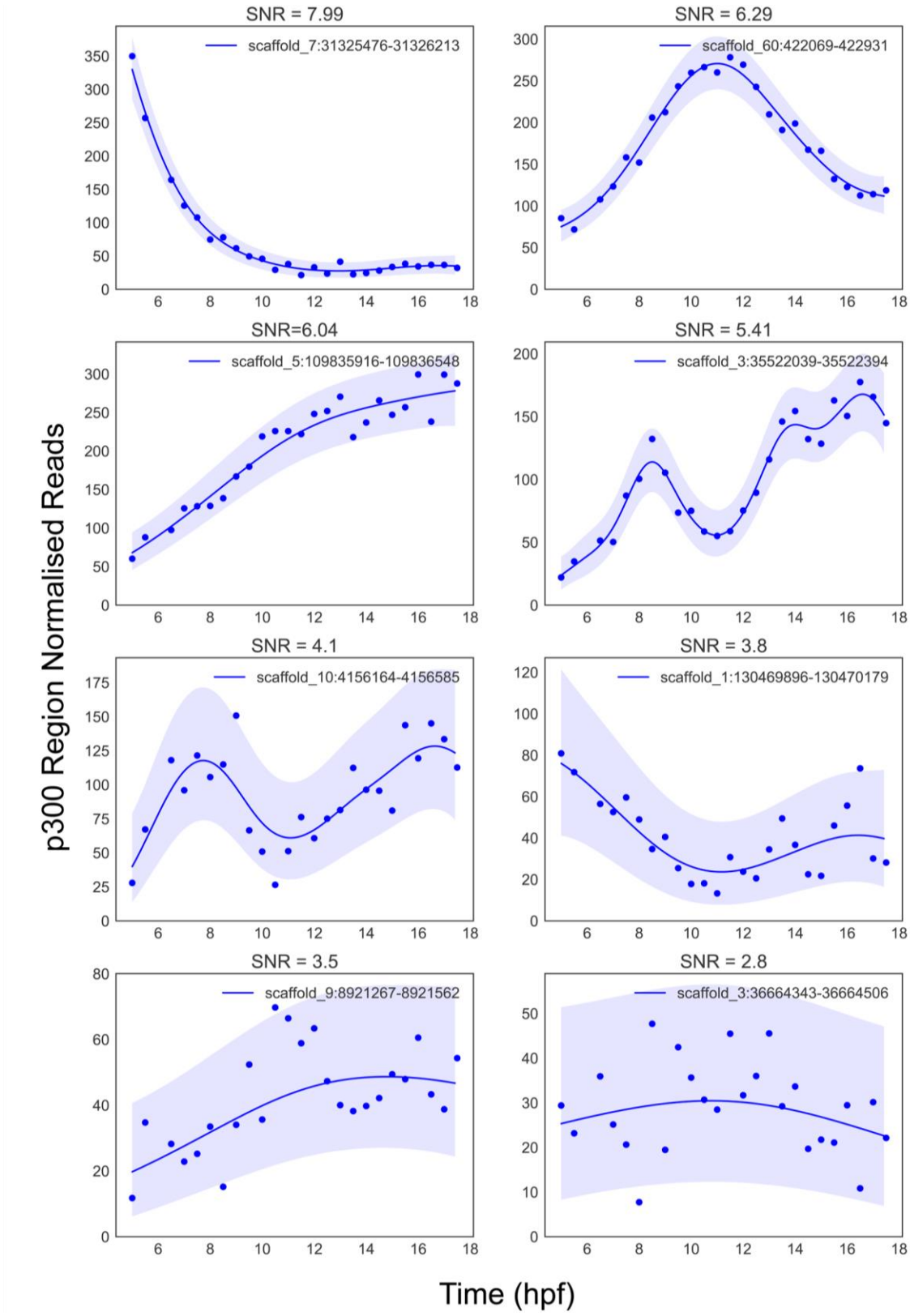


Figure 14 – Examples of p300 binding dynamics and corresponding SNRs.

The y-axis represents the number of normalised p300 reads in a given p300 region. The filtering threshold was set at SNR = 3.8.

3.2.2.7 p300 regions are mainly distal, however there is more promoter-p300 binding than expected for random sequences

To assess where p300 tends to bind in relation to genes, its genomic distribution was calculated. p300 regions were assigned to their closest gene (using HomerTools, as described in Materials and Methods) and given a genomic annotation – intergenic, promoter (for genomic annotation purposes, promoter is defined as the region up to 1kb upstream of the actual TSS), exon or intron. For each p300 region, the distance to the closest TSS was also calculated. In order to compare this data with randomly distributed regions, a random set of regions was created as described in the Materials and Methods.

Figure 15 shows the genomic distribution of p300 regions and compares it with the distribution of a set of random genomic regions.

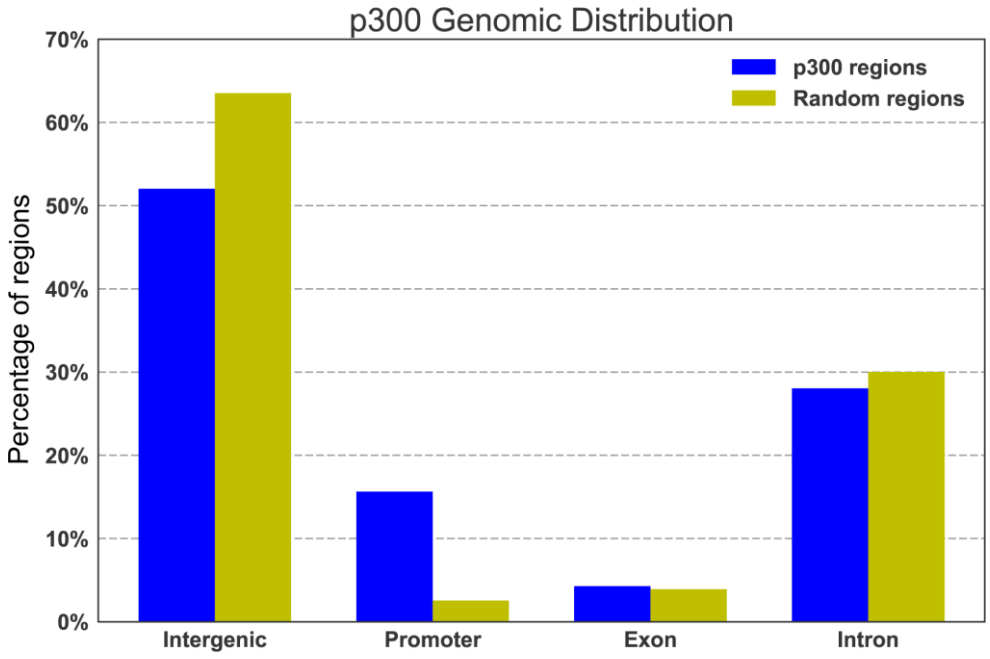


Figure 15 – p300's vs random region's genomic distribution.
Vertical axis represents percentage of p300/random regions with a specific genomic annotation.

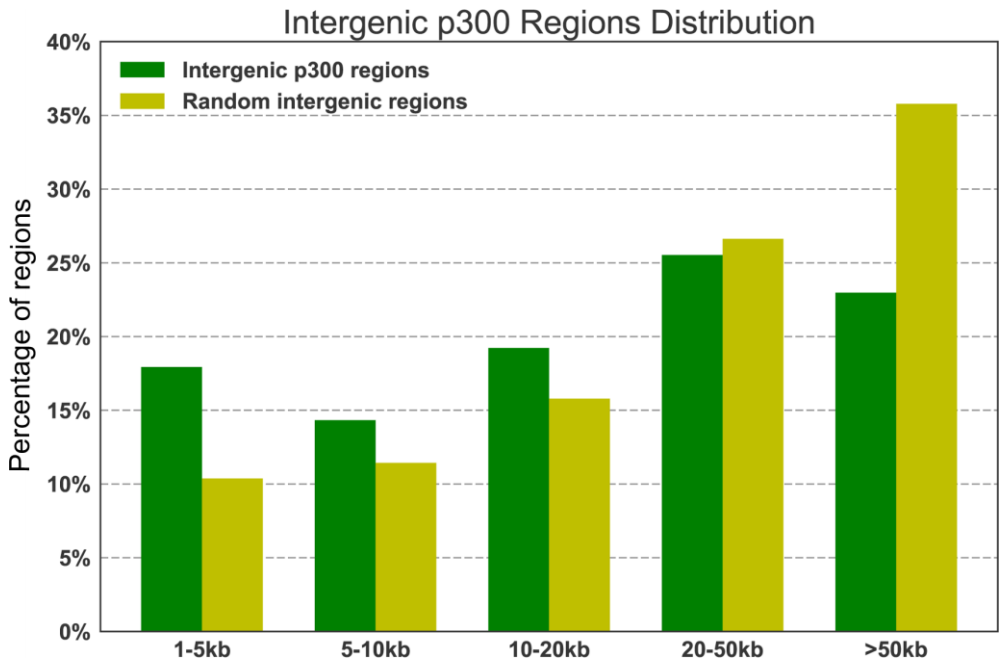


Figure 16 – Intergenic p300 vs intergenic random regions' distance to closest TSS.

Vertical axis represents percentage of intergenic p300/random regions at different distances to the nearest TSS.

Figure 16 shows the distribution of intergenic p300 and random regions, by their distance to the closest TSS. These two figures show that p300 regions tend to be closer to genes than it would be expected for a set of random regions, with Figure 15 showing that p300 regions are 6 times more likely than a random set of regions to be in the gene promoter and Figure 16 showing higher percentages of p300 binding closer to the TSS compared with random regions and the reverse for farther away distances.

From the 12,572 p300 regions, 9,807 are in intergenic or intron regions, therefore, those are the candidate enhancer regions predicted in this study.

3.2.2.8 p300 binding is highly dynamic in early development

In order to determine p300 binding dynamics, the p300 regions were normalised to their maxima and filtered to remove non-dynamic regions - any region with less than a 2-fold change during the time series (excluded 3647 (29%) regions). p300 regions' binding profiles were then clustered by k-means (k = 30), as described in the Materials and Methods, and the resulting clusters were ordered by their average profiles. Figure 17A shows the resulting ordering and Figure 17B shows the constant (< 2-fold change) regions. This figure confirms what Figure 14 had already hinted at, there are very diverse p300 binding dynamics, with some regions having p300 binding only for very short periods. This highlights the importance of having high-resolution time series data; under-sampling would not detect such rapid alterations in binding.

About 38% of p300 regions have their maximum binding at 5 hpf, the first time point, and 28% at 17.5 hpf, the last one. The dynamics of these regions cannot be fully understood because their actual maximum binding may be at any time before or after the time series. 45% of p300 regions have at least one peak in their binding dynamics during the time series (p300 binding profile which increases and then decreases, as the example on Figure 13), with regions being above 95% of maximum binding for, on average, 2.7 hours.

Figure 18 shows the average number of normalised p300 reads in each cluster from Figure 17A. It shows the different types of p300 binding dynamics present in this time series, including some clusters with two distinct stages of high p300 binding.

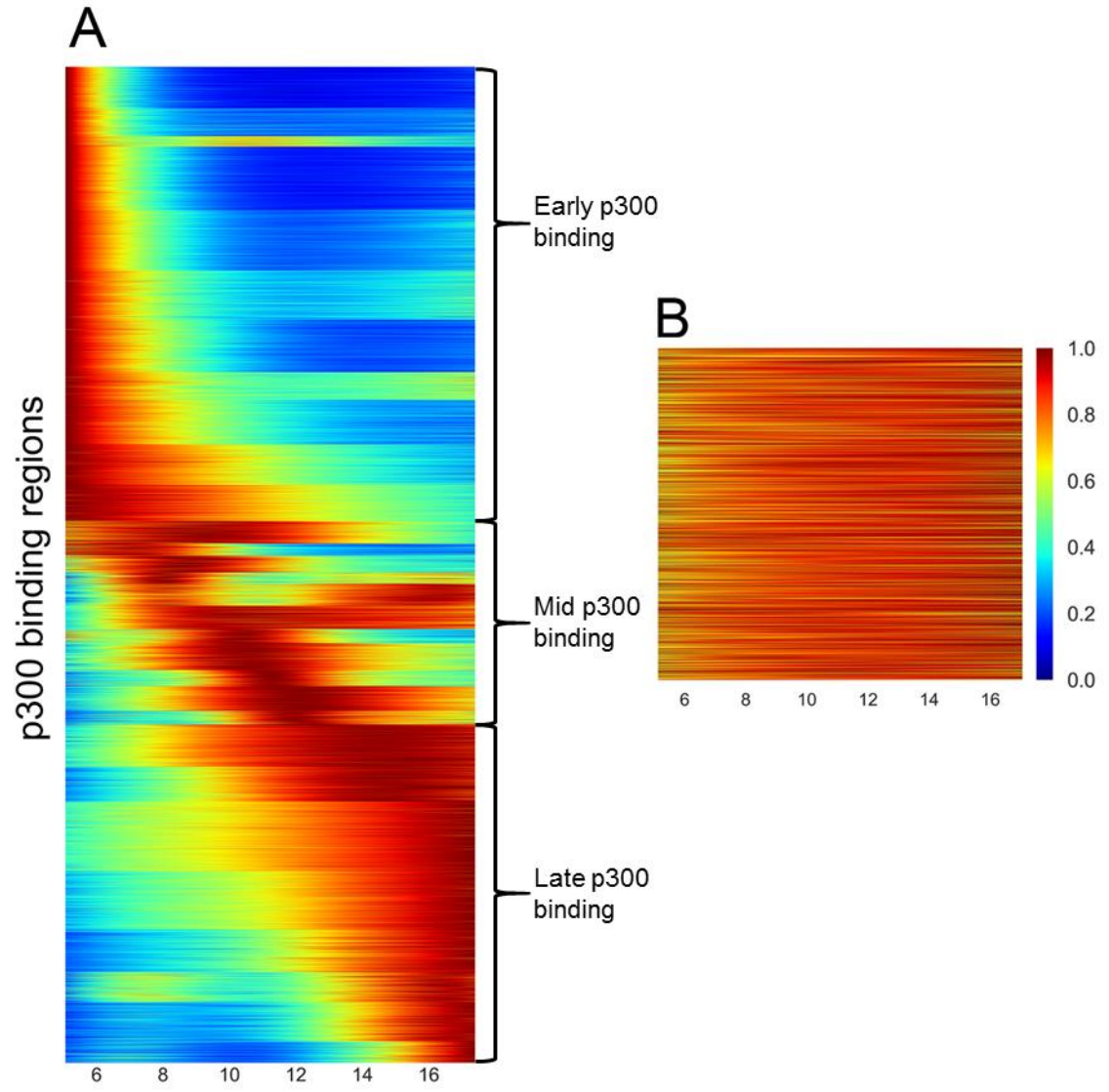


Figure 17 – p300 binding dynamics.

A - Heatmap showing all dynamic p300 regions, normalised by their maxima, kmeans clustered (k = 30) and ordered by cluster maximum time point. p300 regions were divided into early, mid or late p300 binding. B – Heatmap showing all the non-dynamic (< 2-fold change during time series) p300 regions.

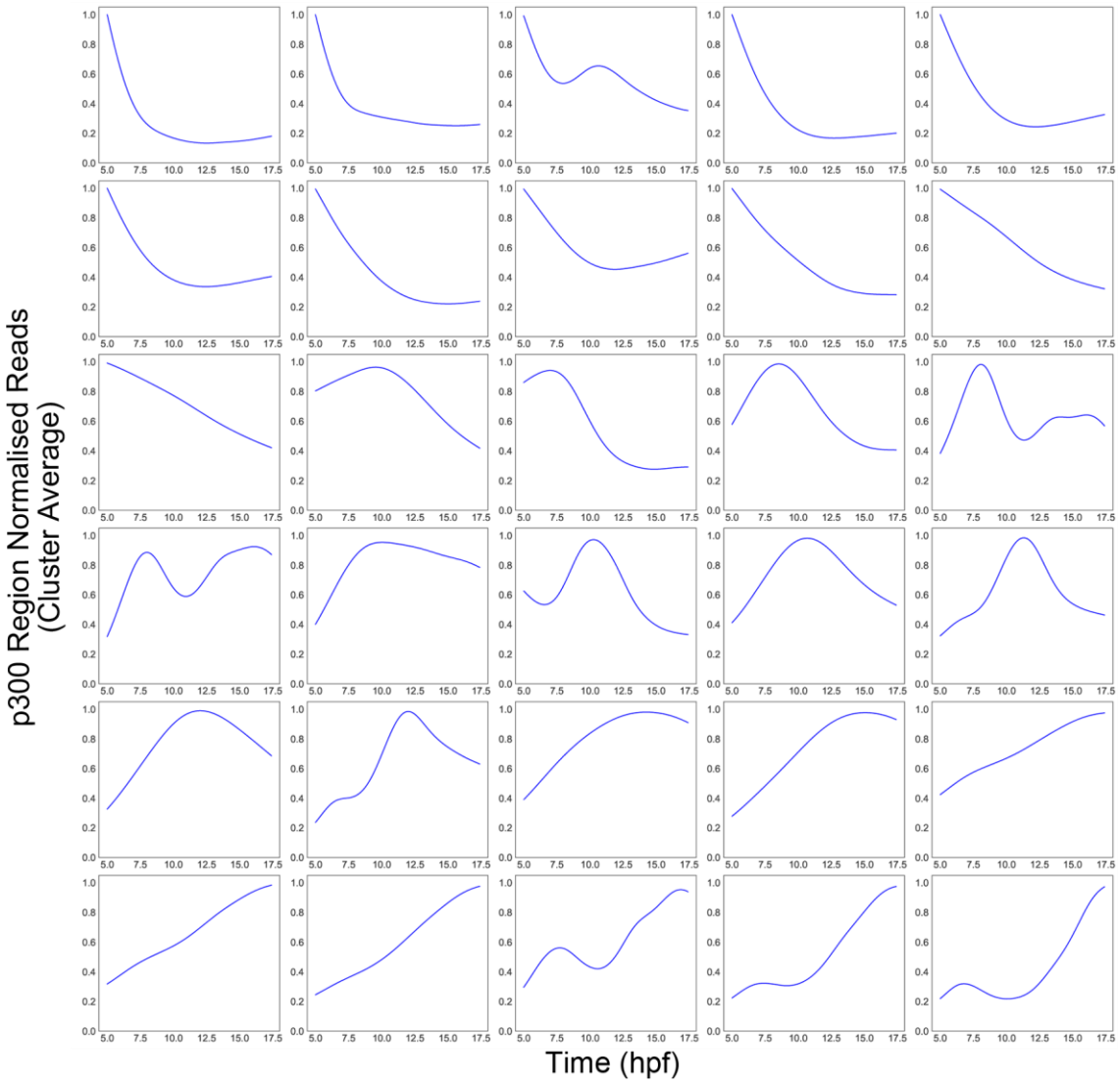


Figure 18 – Cluster p300 dynamics.

Average number of normalised reads in p300 regions per time point in each cluster. Clusters ordered from left to right, then top to bottom.

The gene nearest to each p300 region was determined and GO term analysis was performed using GOstats, focusing on terms for biological processes (GO terms from Owens et al., 2016). Table 8 represents the 10 GO terms with highest enrichment for each group and the corresponding p-value.

GO terms associated with early development, such as gastrulation and Wnt signalling-related terms, are highly enriched in the early p300 binding group. GO terms associated with the development of early tissues and pattern specification are enriched in the mid p300 binding group, while GO terms for specific organ development, particularly kidneys, are enriched in late p300 binding. Finally, GO terms associated with ubiquitous cellular functions, such as transcription, macromolecule biosynthesis and gene regulation are highly associated with genes near p300 regions with constant p300 binding.

	P-value	Gene count	GO term
Early p300 genes (4052 genes)	4.16E-11	172	Cell surface receptor signalling pathway involved in cell-cell signalling
	3.24E-09	150	Cell-cell signalling by wnt
	3.68E-09	149	Wnt signalling pathway
	1.85E-08	117	Gastrulation
	2.35E-08	402	Regulation of multicellular organismal development
	4.13E-08	279	Positive regulation of developmental process
	5.05E-08	667	Anatomical structure morphogenesis
	8.68E-08	413	Locomotion
	1.07E-07	103	Regulation of Wnt signalling pathway
	1.55E-07	269	Tube development
Mid p300 (1804 genes)	3.70E-25	272	Embryo development
	1.26E-22	234	Anatomical structure formation involved in morphogenesis
	1.58E-22	193	Tube development
	5.68E-22	193	Embryonic morphogenesis
	6.98E-22	310	Tissue development
	9.51E-21	244	Epithelium development
	1.63E-20	147	Embryonic organ development
	2.00E-20	186	Embryo development ending in birth or egg hatching
	2.40E-20	170	Pattern specification process
	3.68E-20	181	Chordate embryonic development
Late p300 (2997 genes)	9.24E-09	53	Nephron tubule development
	1.86E-08	56	Renal tubule development
	2.77E-08	58	Nephron epithelium development
	4.00E-08	313	Epithelium development
	6.17E-08	72	Kidney epithelium development
	1.62E-07	68	Nephron development
	1.72E-07	205	Morphogenesis of an epithelium
	2.52E-07	171	Tube morphogenesis
	3.13E-07	46	Renal tubule morphogenesis
	3.68E-07	34	Negative regulation of developmental growth
Constant p300 (3617 genes)	1.81E-22	904	Cellular macromolecule biosynthetic process
	6.59E-22	915	Macromolecule biosynthetic process
	1.70E-21	755	RNA biosynthetic process
	2.42E-21	746	Nucleic acid-templated transcription
	2.42E-21	746	Transcription, DNA-templated
	4.92E-21	841	RNA metabolic process
	1.29E-20	759	Regulation of cellular macromolecule biosynthetic process
	1.44E-20	893	Nucleic acid metabolic process
	2.70E-20	809	Regulation of gene expression
	5.36E-20	923	Gene expression

Table 8 - GO term enrichment in genes near early, mid, late and constant p300 regions

p300 regions were divided into early, mid, late and constant, based on their binding dynamics, and each region was assigned to the closest gene. GO terms associated with genes in each group were analysed and the p-value of a hypergeometric test between genes on each condition and all *Xenopus* genes is presented for each term.

3.2.2.9 p300 is highly correlated to the less dynamic H3K4me1 mark

The p300 dataset was compared to published ChIP-seq data (Hontelez et al., 2015). As previously mentioned, H3K4me1 is usually found at promoters and active or poised enhancers, H3K4me3 and H3K9ac at promoters, H3K36me3 in active gene bodies and H3K27me3, H4K20me3, H3K9me2 and H3K9me3 in heterochromatin or repressed regions.

p300 regions were divided into early, mid, late or constant based on their binding dynamics. The first three groups correspond to the sections on Figure 17A and the constant group corresponds to Figure 17B.

The ChIP-seq read density of each of the published histone modifications and of RNAPII in the set of p300 regions – divided into early, mid, late and constant – were visualised together in order to determine which histone modifications correlate with p300 binding (Figure 19). p300's read density was also plotted for four time points across the time series for reference. Graphs show read count for a 10 kb area centred on the p300 regions. This analysis was performed using DeepTools, as described in Materials and Methods.

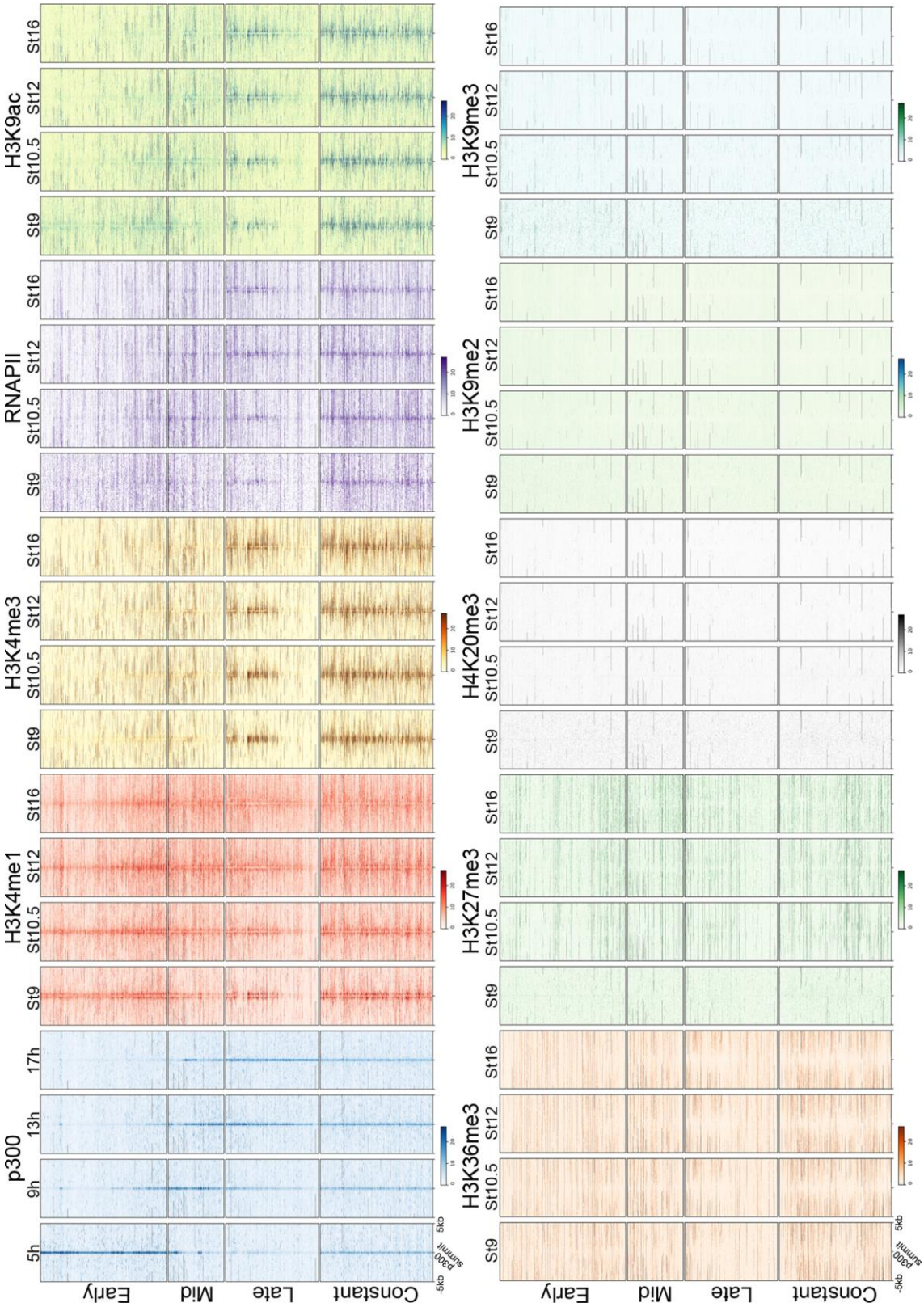

p300
5h
9h
13h
17h
H3K4me1
St9
St10.5
St12
St16
H3K4me3
St9
St10.5
St12
St16
RNAPII
St9
St10.5
St12
St16
H3K9ac
St9
St10.5
St12
St16
H3K36me3
St9
St10.5
St12
St16
H3K27me3
St9
St10.5
St12
St16
H4K20me3
St9
St10.5
St12
St16
H3K9me2
St9
St10.5
St12
St16
H3K9me3
St9
St10.5
St12
St16
Constant
Late
Mid
Early
5kb
p300 summit
-5kb

Figure 19 – Heatmaps of p300 and RNAPII binding and several histone modifications in p300 binding regions.

Regions were ordered by p300 binding dynamics, with the first group being regions with high p300 binding in the beginning of the time series, the second, in the middle, the third in the end of the time series and the last group being regions which p300 binding changes less than 2-fold during the whole time series. p300 data for 5, 9, 13 and 17 hpf are shown for reference. RNAPII and histone modifications data from Hontelez et al., 2015.

As expected, p300's read density is higher at 5 hpf for the early regions, and the read density shifts to the later regions during the time series. Also as expected, the p300 binding in the constant regions (less than 2-fold change) is maintained.

H3K4me1 is present in the majority of p300 regions, which would be expected due to both being present in enhancer regions. H3K4me1 does not appear to be as dynamic as p300, however, the data was independently clustered (Figure 20) in order to determine whether it is indeed less dynamic or if the apparent lower dynamics are due to the temporal profiles not being well correlated. Through hierarchically clustering H3K4me1 in p300 regions, dynamics can be detected, however, p300 binding seems to appear and disappear faster than H3K4me1. This is expected due to this modification being present in both poised and active enhancers.

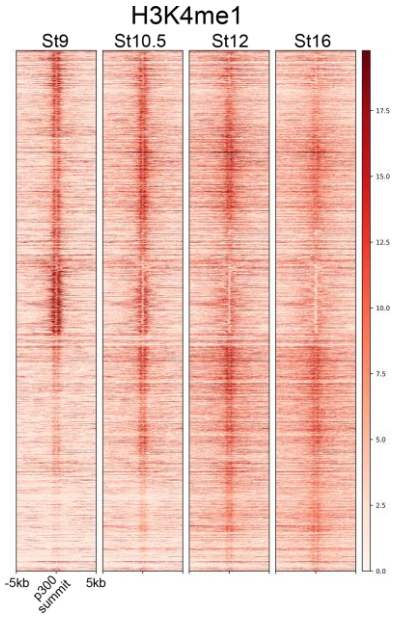


Figure 20 - Heatmap of H3K4me1 in p300 binding regions.

p300 binding regions were hierarchically clustered based on the binding dynamics of H3K4me1. Data from Hontelez et al., 2015.

Overall, H3K4me3, and to a lesser extent H3K9ac and RNAPII, show low levels of colocalisation with p300. However, for the constant p300 regions and for a small cluster of late regions, there seems to be a high H3K4me3, H3K9ac and RNAPII read density. The genomic annotation of those high H3K4me3 regions was analysed and a third of those regions are in promoters, compared to only 16% of the overall p300 regions being in promoters (Figure 21). This would be expected, given that H3K4me3, H3K9ac and RNAPII are present in promoter regions.

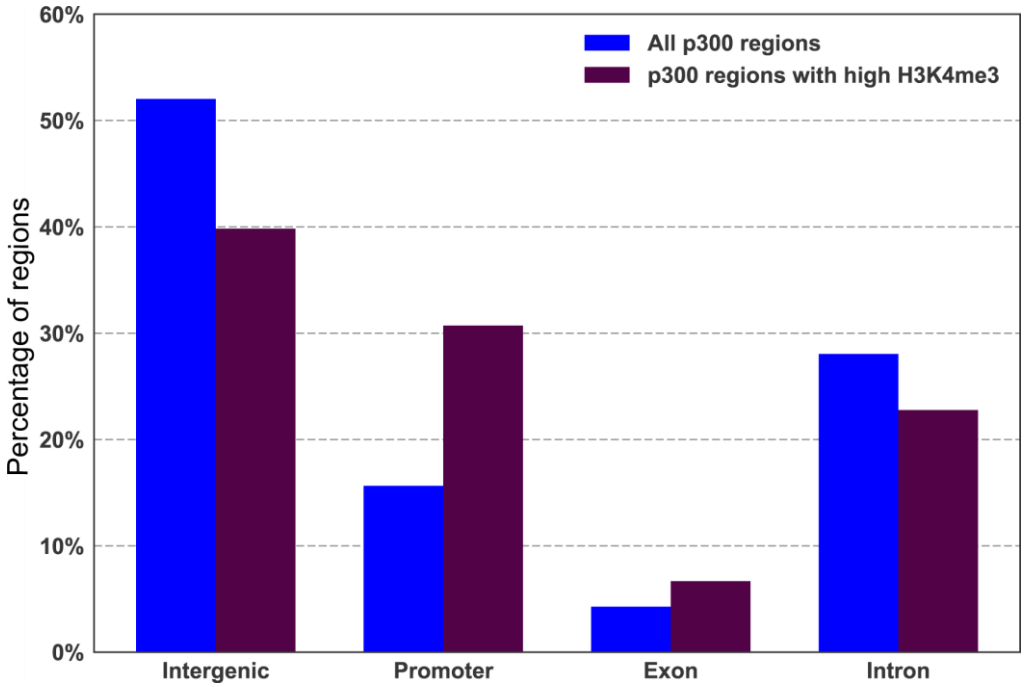


Figure 21 – All p300 regions' vs p300 regions with high H3K4me3's genomic distribution.

Vertical axis represents percentage of all p300 regions/p300 regions with high H3K4me3 with a specific genomic annotation.

H3K36me3, H3K27me3, H4K20me3, H3K9me2 and H3K9me3 do not colocalise with p300 binding, which was also expected, due to these marks typically being present in non-regulatory or repressed regions.

3.2.2.10 p300 binding in promoter and exonic regions is less dynamic

Next, I analysed if p300 regions with a specific genomic annotation (Intergenic, Promoter, Exon or Intron) were more likely to have one of the four different binding dynamics: early, mid, late or constant (divided as described above). The percentage of p300 intergenic regions that have each of the above dynamics was calculated. The same was done for p300 promoter, exonic and intronic regions, as well as for all p300 regions for comparison.

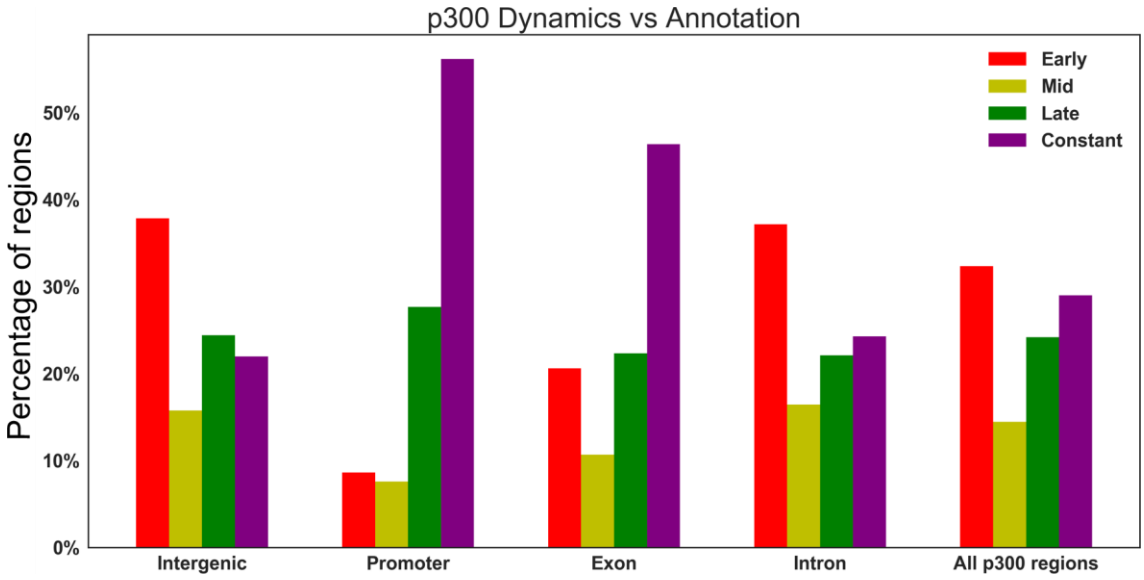


Figure 22 – p300 genomic annotation vs dynamics.

p300 regions were divided based on their genomic annotation (intergenic, promoter, exon, intron or all p300 regions) and then based on their dynamics, if they are active in early, mid or late time points, or if they are not dynamic (constant). Graph shows the percentage of each dynamic in each genomic annotation group. Groups of bars for the same genomic annotation add up to 100%.

Figure 22 shows the results of this analysis, and shows that promoter, and to a lesser extent exonic regions, are much more likely to have constant (non-dynamic) p300 binding. Intergenic and intronic regions have a similar dynamics' distribution as "All p300 regions".

The average fold-change in p300 binding for each genomic annotation group was calculated to determine if p300 binding in promoter regions is indeed less dynamic. Figure 23 shows the distribution of fold change in p300 binding for regions in each of the genomic annotation groups. Promoters do indeed have less dynamic p300 binding than intergenic regions (66% decrease in mean fold change, p-value = 2.73e-235, Mann-Whitney U test). p300 binding at exonic regions is also significantly lower than at intergenic regions (33% decrease, p-value = 1.24e-37, Mann-Whitney U test).

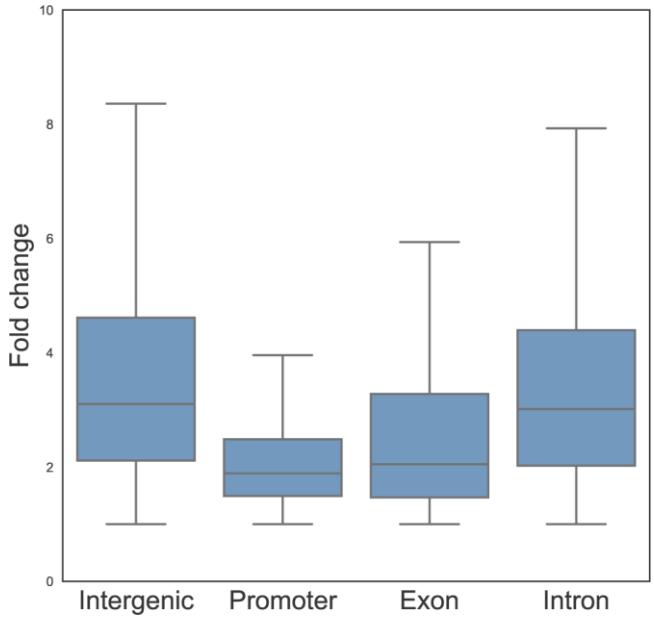


Figure 23 – p300 binding fold change.

Fold change in p300 binding in p300 regions with the different genomic annotations.

In order to determine if different distribution of read densities in promoters and intergenic regions are a confounding factor in the analysis of the dynamic behaviour between the different regions, their distribution was determined (Figure 24).

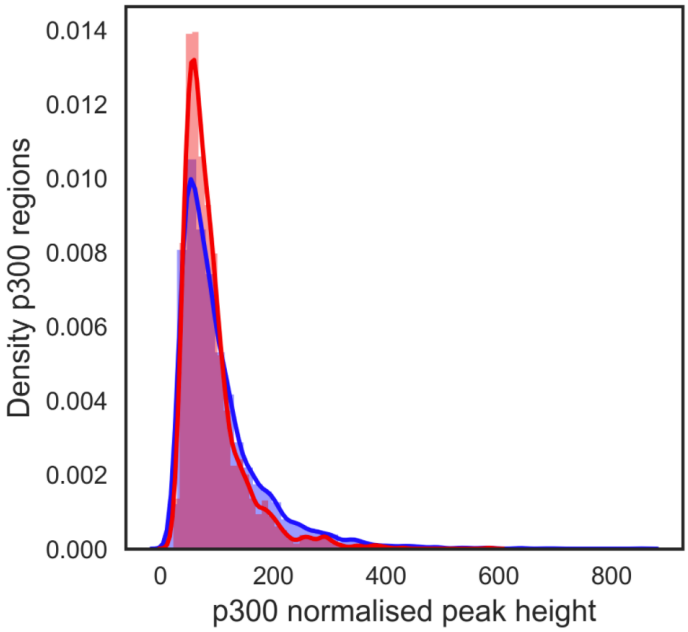


Figure 24 - p300 normalised peak height distribution.
Histogram of p300 normalised peak heights in promoter (red) and intergenic (blue) regions.

To confirm if p300 binding in promoter regions is indeed less dynamic, regions with similar distributions were compared; for this, data was split into high and low p300 binding regions (>150 normalised reads *vs* <100 normalised reads). Regardless of read density, promoter regions do indeed have lower fold change than intergenic regions (65% and 48% decrease in low and high p300 binding regions, respectively. p-value = 1.37e-153 and 3.54e-21, Mann-Whitney U test) (Figure 25).

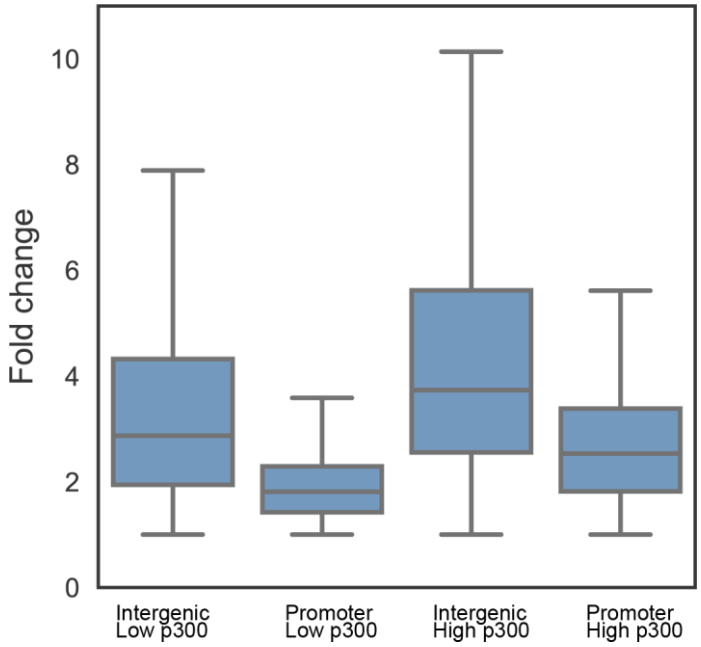


Figure 25 - p300 fold change in regions with low or high p300 binding.

Fold change in p300 binding in intergenic and promoter regions with either low (<100 normalised reads) or high (>150 normalised reads) p300 binding.

Intergenic and promoter p300 regions corresponding to the same nearest gene were paired and their fold changes were analysed. In 74% of the pairs, the p300 binding varied more in the intergenic than in the promoter region, with intergenic regions having, on average, a fold change almost two times higher than the promoter regions.

The expression of the corresponding genes will be analysed in the next chapter, to investigate the relationship between p300 binding at promoters and distal regions with the gene's transcription dynamics.

3.3 Discussion

3.3.1 Pilot time series

In this chapter I described the creation of two p300 ChIP-seq time series. The first of these was a pilot time series which allowed me to test the approach. Despite spanning only 3 hours, it resolved some p300 binding dynamics, and justified the production of a second, longer time series.

The pilot time series confirmed two technical aspects of the approach: firstly, that single-end sequencing is appropriate for ChIP-seq and that little is gained in performing paired-end sequencing; and secondly, for closely spaced time series not all inputs need to be sequenced. These two points translate into worthwhile cost savings, an important consideration when performing high-throughput time series experiments.

3.3.2 Long time series

Next, I performed a long p300 ChIP-seq time series, which showed there is a very high correlation coefficient between adjacent time points and between the two time series. This is important to show that the results obtained are reproducible.

A novel ChIP-seq normalisation method was developed, allowing the comparison of samples with different levels of background signal. This, together with the algorithm adapted from Owens et al., 2016 to generate the curves of best fit for each p300 region, allowed quantitative downstream analysis to be performed.

After filtering out regions with low signal-to-noise ratio, 12,581 p300 regions were found. Importantly, the high-resolution time series allowed the calculation of a temporal signal-to-noise ratio, which allowed the identification of robust p300 binding regions exhibiting consistent temporal dynamics. Additionally, with a lower

temporal resolution I would require replicates of individual time points to predict high confidence p300 regions. Given that I showed that adjacent time points correlate better than biological replicates, they can serve as internal replicates with the application of a framework like Gaussian processes.

Approximately 50% of the p300 regions are intergenic, however, p300 regions tend to be closer to TSSs than expected for a set of random regions, with six times more binding within the 1 kb promoter region. This 50% value is lower than the one reported by Heintzman and colleagues (75% at >2.5 kb) (Heintzman et al., 2007). That study was done in humans, with a genome roughly twice as large as *X.tropicalis* (Hellsten et al., 2010), which may explain the difference.

p300 regions were clustered according to their binding dynamics (early, mid, late or constant binding), which showed that p300 binding in early development is highly dynamic, with more than two thirds of the p300 regions varying more than 2-fold during this time series.

GO analysis was performed on genes near p300 regions in each of the above mentioned groups (early, mid, late and constant p300). Even though not all genes are regulated by their closest enhancer, the enriched GO terms for each category matched well with the developmental processes at each stage. Early p300 regions are present near genes associated with early development and gastrulation, while genes associated with tissue patterning and organ development are near mid and late p300 regions, respectively. Most interestingly, genes involved in ubiquitous cellular functions were highly enriched near p300 regions which have constant binding. These results reinforce that p300 binding is regulating expression of different genes at different stages of development.

When comparing the generated data with previously published data on histone modifications, I found good agreement with the literature. p300 and H3K4me1 colocalise well, however p300 appears more dynamic than this histone modification. Overall, p300 and H3K4me3/H3K9ac/RNAPII do not correlate well, except for the constant p300 regions, and a subset of the late p300 regions. These regions were analysed and, unsurprisingly, a much higher percentage of these are in promoter regions, when compared to the overall p300 binding. This was also expected due to these three marks being present in active promoter regions.

Finally, 56% of p300 binding to promoters is not dynamic, changing less than 2-fold during the 12.5 hours of this time series, compared to only 21% of p300 binding in intergenic regions being constant. This indicates that p300 binding is much less dynamic in promoters than in distal enhancers. The average fold change in p300 binding in the different genomic annotation groups was calculated and the p300 binding at promoter regions does indeed vary significantly less than at intergenic regions. The analysis of intergenic and promoter regions close to the same gene showed that intergenic p300 binding varies almost twice as much as the p300 binding in the nearest gene's promoter. These values are likely to be underestimated, due to the existence of unannotated promoters in the *X. tropicalis* genome. As mentioned in the Introduction, p300 is implicated in enhancer-promoter looping. If one promoter interacted with only one enhancer, it would be expected that they would have very similar p300 binding dynamics. However, different enhancers regulate gene expression in different cell types and at different developmental stages, thus, promoters may be involved in looping, with different enhancers, for longer periods of time. The observation that p300 binding in distal candidate enhancers is more dynamic than in promoter regions is then a

reinforcement of the model of a promoter binding different enhancers in different tissues at specific time periods.

In this chapter I described the creation of a set of 9,807 candidate enhancer regions in early *X. tropicalis* development, showed that p300 binding is extremely dynamic and that p300 binding in promoter regions is less dynamic than in distal candidate enhancers. In the next chapter I will describe the analysis performed in order to understand how p300 binding correlates with gene transcription and how it may be used to predict enhancer-gene pairs.

Chapter 4. p300 and Gene Transcription

4.1 Introduction

The aim of the analysis described in this chapter was to investigate the temporal relationship between p300 binding and gene transcription, in order to detect potential correlations and to assess if they can be used to predict enhancer-gene pairings.

p300 may regulate transcription through various mechanisms, with one of the most prominent being a proposed role in enhancer-promoter looping (Kim and Kim, 2013, Guo et al., 2016). If p300 is involved in activating transcription, and not only establishing a poised transcriptional state which lasts several hours until transcription actually starts, the dynamics of p300 binding should correlate with transcription dynamics.

Based on this, I set out to answer the following questions:

1. Are active genes (genes with increasing number of transcripts) more likely to have p300 binding at their promoter and/or nearby candidate regions? If p300 binding is indeed involved in transcription regulation, it would be expected that active genes are more likely to have nearby p300 binding than inactive genes.
2. Do p300 binding dynamics at promoters correlate with the binding dynamics at nearby enhancers? If p300 is involved in enhancer-promoter looping, it would be expected that the binding dynamics in an enhancer and corresponding promoter to be similar, as the same molecule is in proximity to both regions.

3. Do p300 binding dynamics in a gene promoter correlate with the transcription rate of that gene? If p300-mediated looping leads to transcription initiation and the loop is maintained during transcription, it would be expected that the more cells have p300 binding in a given promoter, the more that gene would be expressed.
4. Can enhancer-gene pairs be predicted based on the two previous hypotheses and what can be said about conventional enhancer-gene pairing approaches, such as pairing an enhancer to the closest gene, in light of this data?

4.2 Active genes are more likely to have p300 binding nearby

I first set out to assess if active genes are more likely than inactive ones to have p300 binding in their promoter regions and/or in nearby candidate enhancers.

A list of 1677 genes activated between 6 and 19 hpf and their fold change (between activation and their peak) was generated by Michael Gilchrist. Briefly, a gene is described as active if its expression levels rise sharply over consecutive time points (method described in Collart et al., 2014).

These genes were sorted by their fold change and divided into three groups: top 33%, middle 33% and bottom 33% (559 genes in each group) of active genes. The percentage of genes within each category, as well as in all inactive genes (all genes not considered active by the above method, which may include some genes with low activation - 26,713 genes), with a p300 region in the 100 bp upstream or downstream of the TSS (hereafter referred to as *proximal promoter*) or in the 20 kb surrounding region was determined. The 200 bp region around the promoter was removed from the 20 kb region. As described in the previous chapter, a p300 region is any region detected by the peak caller in any sample and merged, followed by SNR filtering.

Table 9 summarises the results. Genes in the top 33% set are almost 6 times more likely to have a p300 region in the proximal promoter than inactive genes (p-value = 8.56e-65, Fisher Exact test), and almost 2 times as likely as the middle and bottom 33% active genes (p-value = 3.34e-08 and 2.33e-06, respectively, Fisher Exact test). Regarding the 20 kb surrounding region, the top 33% active genes are almost 4 times more likely than inactive genes to have a p300 region within this region (p-value = 2.01e-102, Fisher Exact test), and 1.4

times more likely than the bottom and middle 33% of active genes (p-value = 2.46e-07 and 8.93e-09, respectively, Fisher Exact test).

	Proximal Promoter	20kb region
Inactive Genes	1385 (5%)	4208 (16%)
Bottom 33% active genes	97 (17%)	237 (42%)
Middle 33% active genes	87 (16%)	227 (41%)
Top 33% active genes	155 (28%)	327 (59%)

Table 9 – p300 at proximal promoters or nearby genes with different transcriptional states.

Percentage of inactive (26,713 genes), bottom (559 genes), middle (559 genes) and top (559 genes) 33% of active genes that have p300 binding in their proximal promoter (±100bp) or in the 20 kb surrounding region. Active genes were divided based on their expression fold change.

To determine if there was more p300 binding (i.e. more normalised p300 reads) in proximal promoters of active genes, the maximum p300 binding during the time series in the regions in the four categories was determined (Figure 26). Proximal promoters of the top 33% active genes have a maximum p300 binding 45% higher than proximal promoters of inactive genes (p-value = 9.22e-12, Mann-Whitney U test). The difference between the three sets of active genes is small, with top vs middle 33% not being statistically significant and top vs bottom 33% only representing a 6% increase in maximum p300 binding (p-value = 0.0045, Mann-Whitney U test).

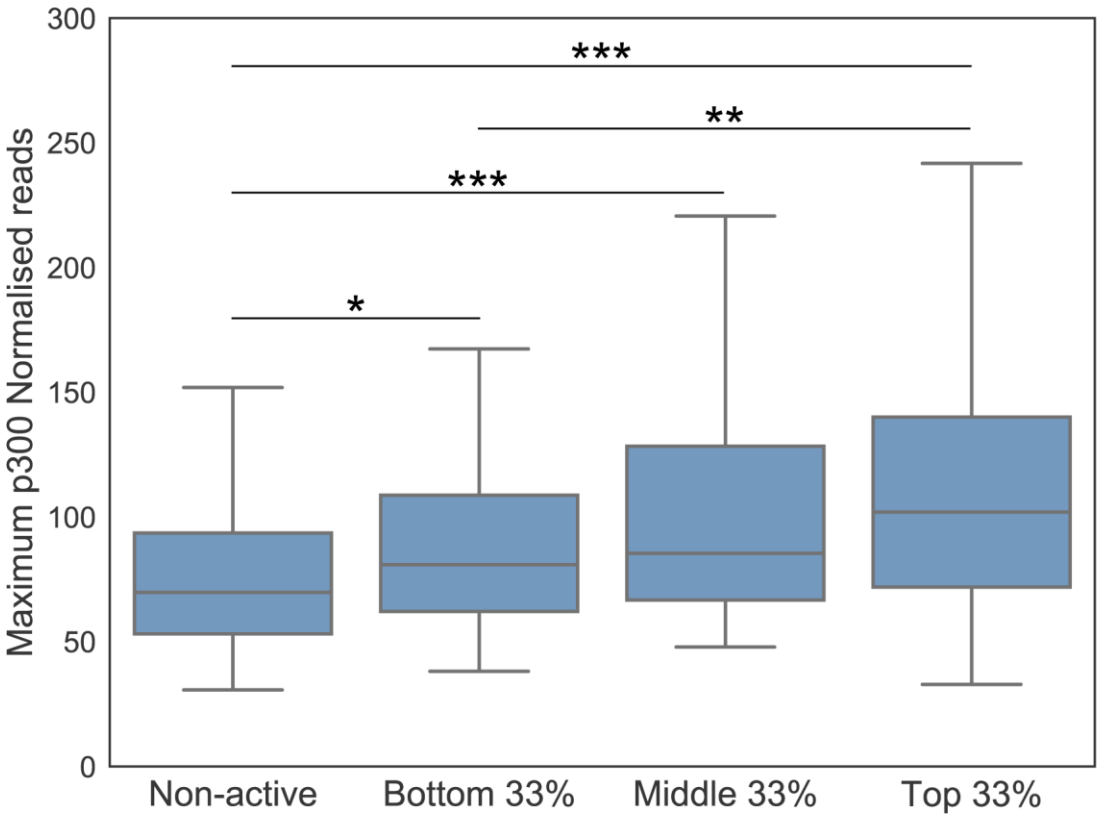


Figure 26 – p300 occupancy at proximal promoters of genes with different transcriptional states.

Boxplots of maximum p300 binding at proximal promoter regions of inactive, bottom, middle and top 33% active genes. *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001.

4.3 p300 dynamics and nearest gene's expression

Pairing genes with their corresponding enhancers is a challenge, with numerous genes known to be regulated by enhancers at extremely long distances (e.g. *shh* enhancer 1 Mb away from the gene promoter (Lettice et al., 2003)). In this section, I describe an initial analysis by pairing p300 regions with their closest gene, to determine if there is any overall correlation.

Genes were divided into three groups: genes with early (4052 p300 regions), mid (1804 p300 regions) or late (2997 p300 regions) p300 binding (based on the p300 dynamics calculated in the previous chapter – 3.2.2.8 – p300 binding is highly dynamic in early development).

I set out to determine if a gene close to a region which is bound by p300 during early time points is transcribed at early time points, and the same for genes near mid and late p300 regions. To analyse this, I first needed to calculate the net transcription rate for each gene from the Owens et al., 2016 data.

4.3.1 Net transcription rate

As mentioned in this thesis Introduction, Owens and colleagues developed a method to calculate the absolute normalisation of the RNA-seq data (Owens et al., 2016). This absolute normalised data then allows the calculation of net transcription rates, which represent the rate of active transcription minus the rate of transcript degradation (expressed in transcripts/min/embryo) using:

$$net\ transcription\ rate = \frac{dg(t)}{dt}$$

with *g(t)* describing the expression level of a gene at time *t* (calculated using the Gaussian process method described in the previous chapter). Figure 27 shows an example of a gene's expression (*nodal3.1*) and its net transcription rate.

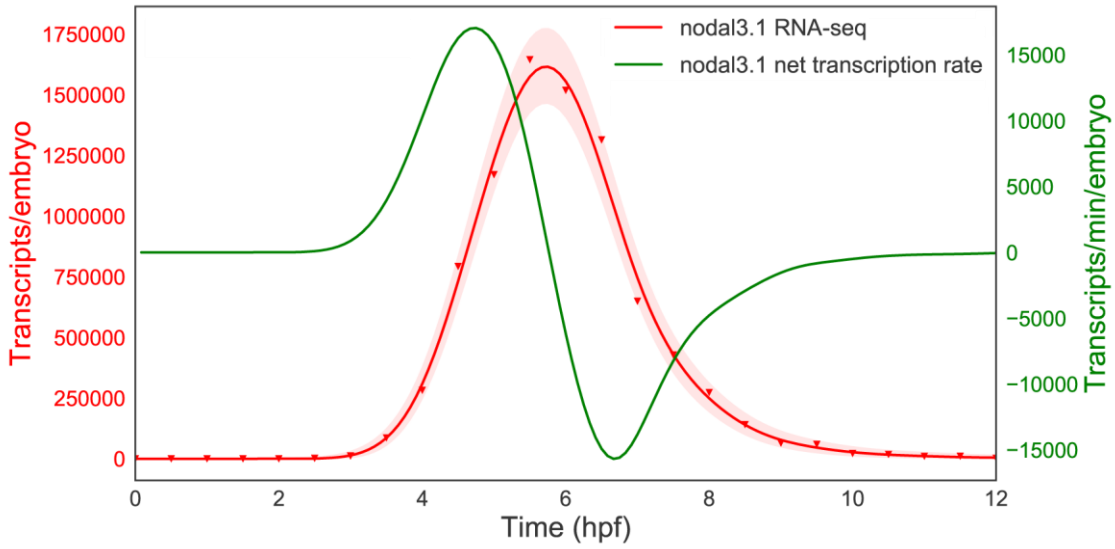


Figure 27 – Net transcription rate.
Nodal3.1 absolute transcript levels (red) and its net transcription rate (green).

The RNA-seq experiment described in Owens et al., 2016 was performed in a clutch with a division time during the cleavage stages of 20 minutes, while the long p300 ChIP-seq time series' embryos had a division time of 18 minutes. To perform comparative analysis between these two time series, the times on the p300 data were adjusted by multiplying each time by 20/18 (a procedure used and validated in Collart et al, 2014 to align RNA-seq time series developing at different rates), e.g. the first time point was adjusted from 5 hpf to 5.56 hpf and the final time point from 17.5 hpf to 19.44 hpf.

4.3.2 Active genes near early p300 binding are transcribed early on in the time series

In order to determine if the presence of early, mid or late p300 binding determined the timing of the closest gene's expression, I calculated the average normalised transcript level and the average net transcription rate, over the p300 ChIP-seq time series period, of the genes closest to the p300 regions in each group of p300 binding dynamics – early, mid or late. This was done considering 1) top 33% active genes (186 early, 151 mid and 144 late); 2) middle 33% active genes (111 early, 107 mid and 112 late); 3) bottom 33% active genes (119 early, 75 mid and 129 late); 4) inactive genes (2251 early, 884 mid and 1692 late). The number of genes in each group is lower than the total number of active/inactive genes as some are not the closest gene to any p300 region.

The results are summarised on Figure 28. It first shows the average p300 dynamics in the early, mid and late groups, for comparison. The top 33% active genes near p300 regions with early dynamics reach their maximum (plateau) earlier than genes near mid and late p300. This becomes more obvious when

analysing the net transcription rate: genes near early p300 regions have their maximum net transcription rate at around 6 hpf, while for genes near mid p300 regions it peaks at around 10 hpf and for genes near late p300 regions at 15 hpf. This difference is less pronounced when analysing the middle 33% active genes and it is almost non-existent in the bottom 33% active genes. There is no significant difference for inactive genes, except that genes near late p300 regions have a higher net transcription rate at later time points. A possibility is that a subset of these genes are activated just after the end of the p300 time series and their net transcription rate is already increasing during those time points.

Even with a naïve enhancer-gene pairing method, based only on genomic distance, it is possible to observe a relationship between p300 binding dynamics and gene expression, particularly for highly active genes.

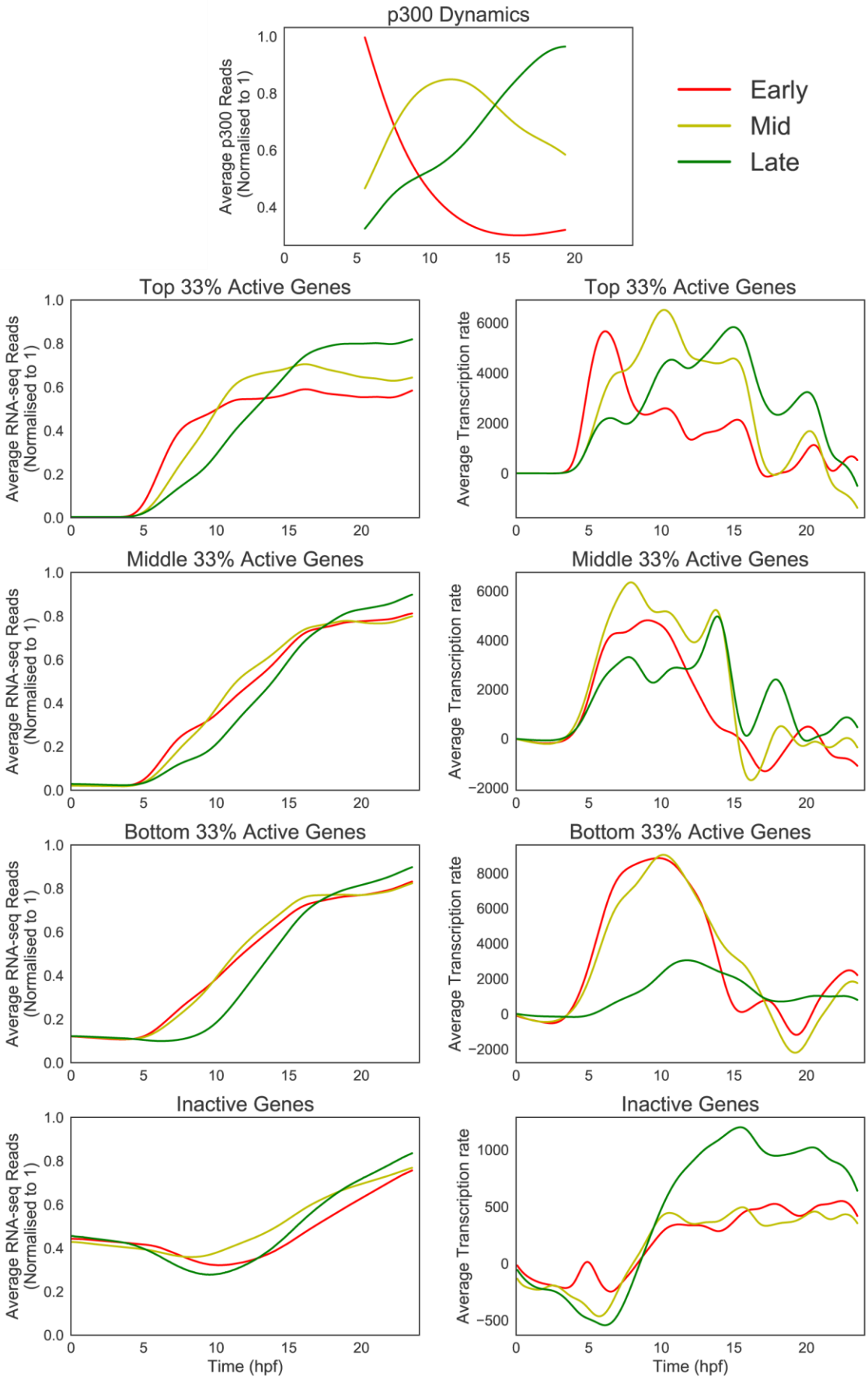

p300 Dynamics
Average p300 Reads (Normalised to 1)
Early
Mid
Late
Top 33% Active Genes
Average RNA-seq Reads (Normalised to 1)
Top 33% Active Genes
Average Transcription rate
Middle 33% Active Genes
Average RNA-seq Reads (Normalised to 1)
Middle 33% Active Genes
Average Transcription rate
Bottom 33% Active Genes
Average RNA-seq Reads (Normalised to 1)
Bottom 33% Active Genes
Average Transcription rate
Inactive Genes
Average RNA-seq Reads (Normalised to 1)
Time (hpf)
Inactive Genes
Average Transcription rate
Time (hpf)

Figure 28 – p300 and transcription dynamics.

Average number of p300 normalised reads in early, mid and late p300 regions (normalised to 1). Average transcript levels (normalised to 1) and net transcription rates for top, middle and bottom 33% active genes, as well as for inactive genes, divided based on the dynamics of p300 binding to nearby candidate enhancers.

4.4 Promoter vs distal p300

p300 is known to be involved in enhancer-promoter looping (Kim and Kim, 2013, Guo et al., 2016), thus, I hypothesised that the dynamics of p300 binding to a promoter and to its corresponding enhancers should be similar, given that for an individual loop, the same p300 molecule will be localised close to both the enhancer and the promoter for the duration of the loop.

In order to determine if p300 binding dynamics at proximal promoters and nearby candidate enhancers are similar, the Euclidean distance between the two curves was calculated. Euclidean distance (ED) represents the distance between the points of two curves, represented by the following equation:

$$||u - v|| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}$$

with *u* and *v* representing each curve (normalised to their maxima) and *n* representing the number of points in each curve. The lower the ED, the more similar the two curves are.

The ED between the p300 binding at proximal promoter regions and in every non-promoter p300 region in the surrounding 20 kb region (*nearby p300 regions*) was calculated, as well as between each proximal promoter region and random p300 regions (in a different chromosome) for comparison. As previously mentioned, enhancer-promoter looping can occur at much larger distances, however, using a 20 kb region allows for an initial, less noisy, analysis.

The mean ED between p300 binding at proximal promoters of active genes and at nearby p300 regions is 3.51, 17% lower than the distance between random pairs (p-value = 3.08e-14, Mann-Whitney U test) (Figure 29).

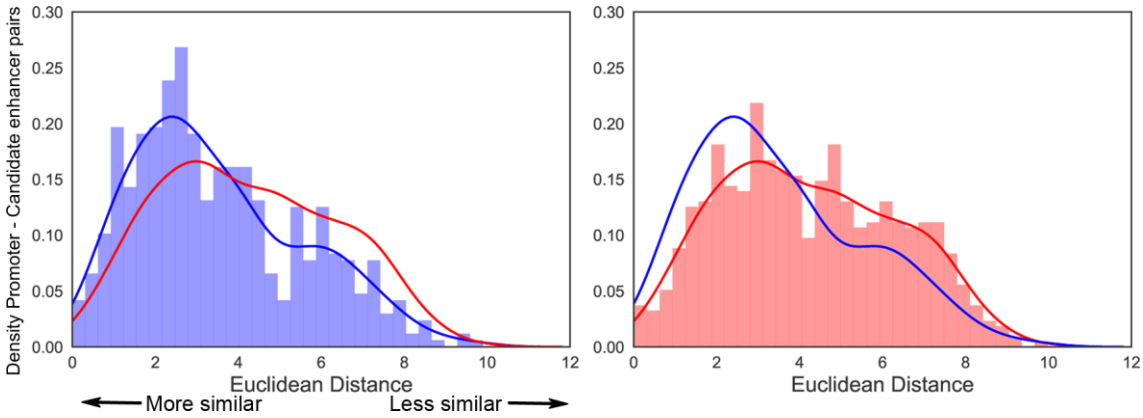


Figure 29 – Euclidean distances between p300 binding at proximal promoters and nearby p300 regions.

Histograms and distribution curves representing the EDs between p300 binding at proximal promoters of active genes vs nearby candidate enhancers (blue) or random candidate enhancers from a different chromosome (red).

These results show that p300 binding dynamics at proximal promoter regions are significantly more similar to the binding dynamics in nearby p300 regions than to random candidate enhancers (on a different chromosome). Some of the random pairings also have low ED, thus they have similar dynamics. This will be addressed in this chapter's discussion.

Figure 30 shows an example of p300 binding at a proximal promoter region (green) and at three nearby candidate enhancers (in blue) with low ED, demonstrating how similar the dynamics are. It is also interesting to note how the four regions show the same p300 binding dynamics.

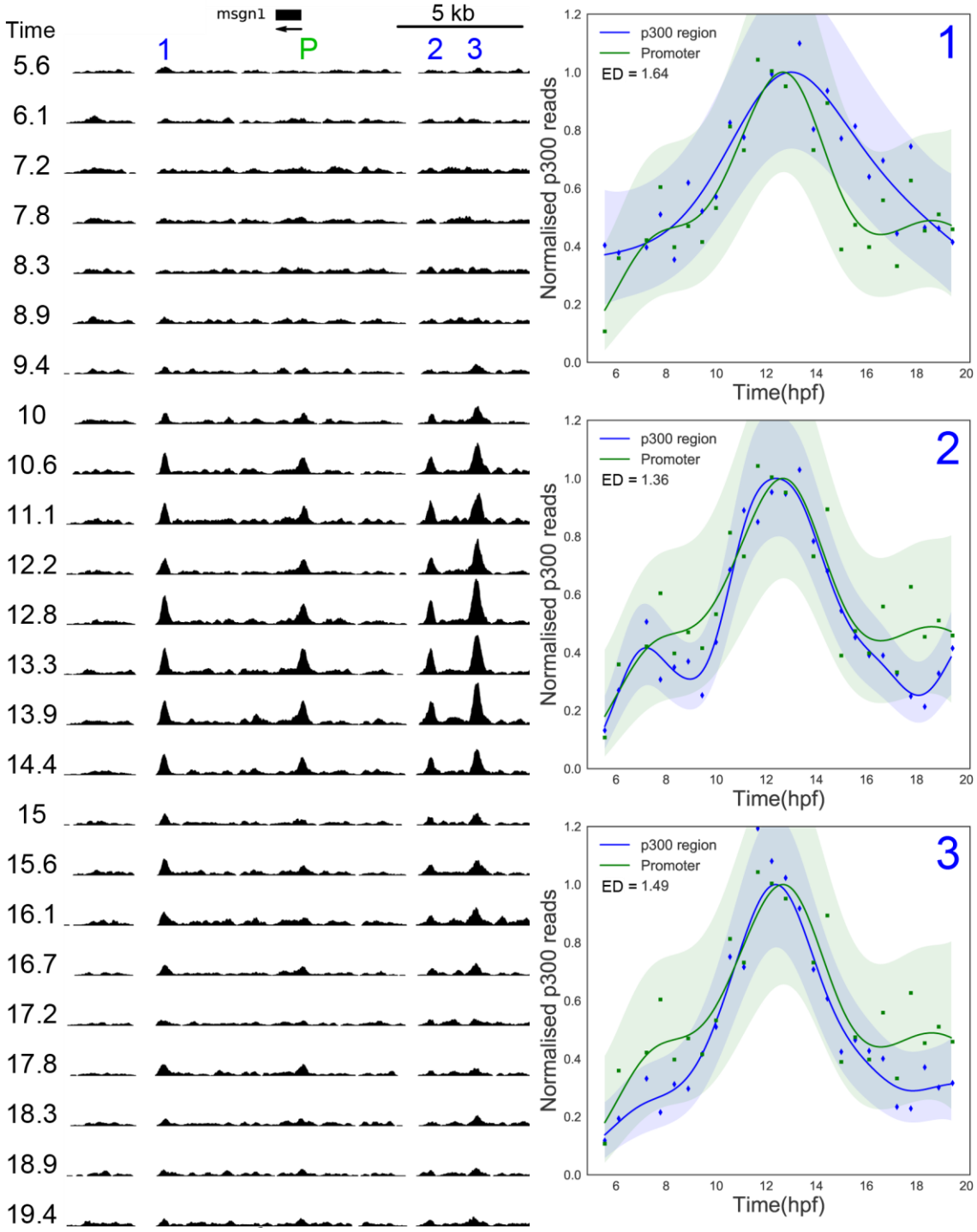


Figure 30 – Example of correlated p300 dynamics.

Example of p300 binding at a proximal promoter (P – green) and three nearby candidate enhancers (1, 2, 3 – blue). Genome browser view of the four regions and graphs of the p300 binding dynamics of each pair, with the corresponding ED.

4.5 p300 binding at promoters, transcript levels and net transcription rate

In the previous section I have shown that p300 binding at proximal promoters correlates more closely with nearby candidate enhancers than with random p300 regions. Whereas enhancers do not always regulate the closest gene, the p300 binding in a gene's promoter is likely to be involved in regulating that gene's transcription. This allows the analysis of whether and how p300 binding correlates with transcription, without the confounding issues associated with pairing enhancers and genes.

ChIP-seq measures how many cells in the embryo have a particular protein bound to a DNA sequence relative to all other DNA sequences. I hypothesised that the more cells have p300 involved in a specific enhancer-promoter looping, the more its target gene is expressed and therefore, the gene's transcription rate should correlate with the p300 binding dynamics at the gene's promoter. For sharply increasing active genes, the net transcription rate, described in section 4.3.1 - Net transcription rate, should provide a reasonable approximation for the actual transcription rate. Moreover, I calculated a scale normalised Euclidean Distance between p300 regions and genes, asking whether to what extent the shapes of the curves match; this should partly abrogate the effect of not accounting for transcript degradation.

To assess how similar p300 and transcription dynamics are, I calculated the ED between p300 binding at proximal promoters and both that gene's absolute transcript levels over time and net transcription rate.

p300 binding at active genes' proximal promoters correlates better with genes' transcript levels than expected at random (mean ED = 4.74, 22% lower than

for random pairs, p-value = 3.30e-28, Mann-Whitney U test) (Figure 31A), however the same is not the case for the genes' net transcription rate (p-value = 0.075, Mann-Whitney U test) (Figure 31B). Furthermore, proximal promoter vs gene transcript levels ED is 54% lower than proximal promoter vs net transcription rate ED (mean ED = 4.74 vs 10.30, respectively, p-value = 2.91e-107, Mann-Whitney U test), demonstrating that proximal promoter p300 binding profiles correlates better with gene transcript levels than with net transcription rates.

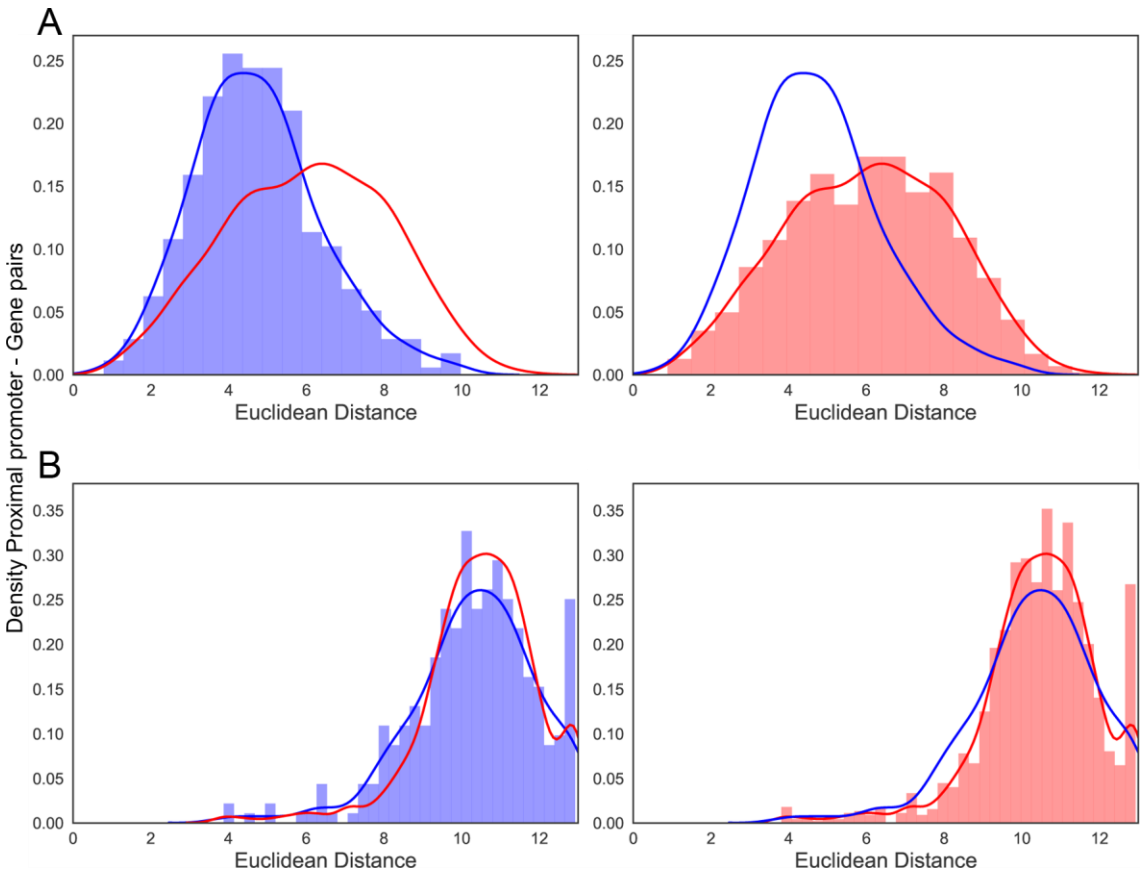


Figure 31 – Euclidean distances between p300 binding at proximal promoters and gene transcription levels/net transcription rate.

Histograms and distribution curves representing the ED between p300 binding at proximal promoters vs A – Gene's absolute transcript levels over time (blue), or B – Gene's net transcription rate (blue). Comparison with p300 binding at random proximal promoters in a different chromosome are represented in red.

These results fail to support the initial hypothesis that p300 binding at proximal promoters should correlate with the net transcription rate of the corresponding gene.

Figure 32 shows examples of p300 binding in proximal promoters (green) and the corresponding gene's transcript levels over time (red) with low ED, showing how similar their profiles are.

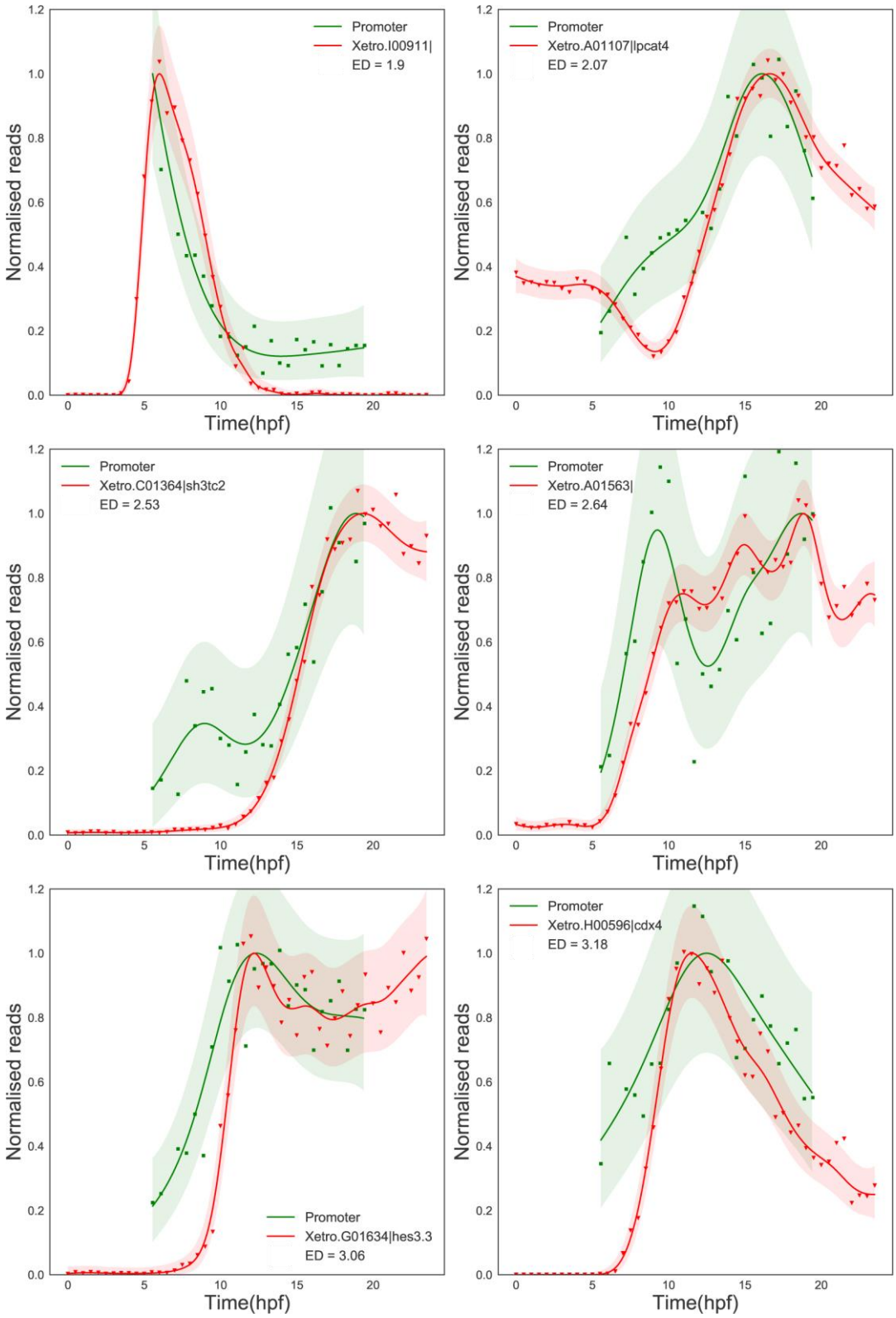


Figure 32 – Highly correlated p300 binding at proximal promoter and gene transcript levels.

Examples of p300 binding at a proximal promoter (green), the corresponding gene's transcript levels (red) and the pair's ED.

4.6 Candidate enhancers – gene pairing

In the previous two sections it was shown that there is correlation between both the dynamics of p300 binding at proximal promoters and nearby candidate enhancers and the dynamics of p300 binding at proximal promoters and the corresponding gene's transcript levels over time. The promoter-candidate enhancer analysis alone cannot predict candidate enhancer-gene pairings, as there may be cases where that correlation is high but that the p300 promoter-gene correlation is not. In such cases, it is possible that those genes are either not regulated by p300-mediated looping or that other events must occur before transcription begins.

I hypothesised that enhancer-gene pairs could be predicted based on both analyses: If the p300 binding at a candidate enhancer correlates well with the binding at a gene promoter *and* the binding at a gene promoter correlates well with the gene's transcript levels, I would predict that that candidate enhancer has a higher likelihood of regulating that gene.

To this end, I developed a method to predict candidate enhancer-gene pairs, by calculating the Sum of Euclidean Distances (SED). The EDs calculated in the two previous sections (candidate enhancer-proximal promoter and proximal promoter-gene) were added, to reach a SED score for each candidate enhancer-gene pair in a 200 kb region (distance can be adjusted). In summary:

$$Sum\ Euclidean\ Distances(SED)_{E-G} = ED_{P-E} + ED_{P-G}$$

with *ED* representing Euclidean distance, *P* the p300 binding at the proximal promoter of active genes, *E* the p300 binding at candidate enhancers and *G* the corresponding gene's transcript levels. Lower SED scores indicate higher correlation.

For each active gene with p300 binding in its proximal promoter (311 genes), the candidate enhancer with the lowest SED score was predicted as the gene's corresponding enhancer (7.1 – Appendix 1 – Predicted enhancer-gene pairs). This method can also be used to predict several enhancers for the same gene and order them by their score.

From these 311 active genes with p300 binding in their proximal promoter, 242 have three or more candidate enhancer regions in the 200 kb surrounding region. The closest candidate enhancer is the best match only for 25% of these genes, while 28% of the genes have their best match on the 2nd or 3rd closest candidate enhancer. This means that for almost half of the genes (47%) their best candidate enhancer match is the 4th closest or further.

As an example of this, *Xetro.A01863* and the six candidate enhancer regions in a 170 kb region are shown on Figure 33, as well as the gene's RNA abundance, the p300 binding dynamics in its proximal promoter and in the six candidate enhancers. The most well correlated region is the 5th one (ordered by increasing distance), located at 43 kb, in an intron of a nearby gene.

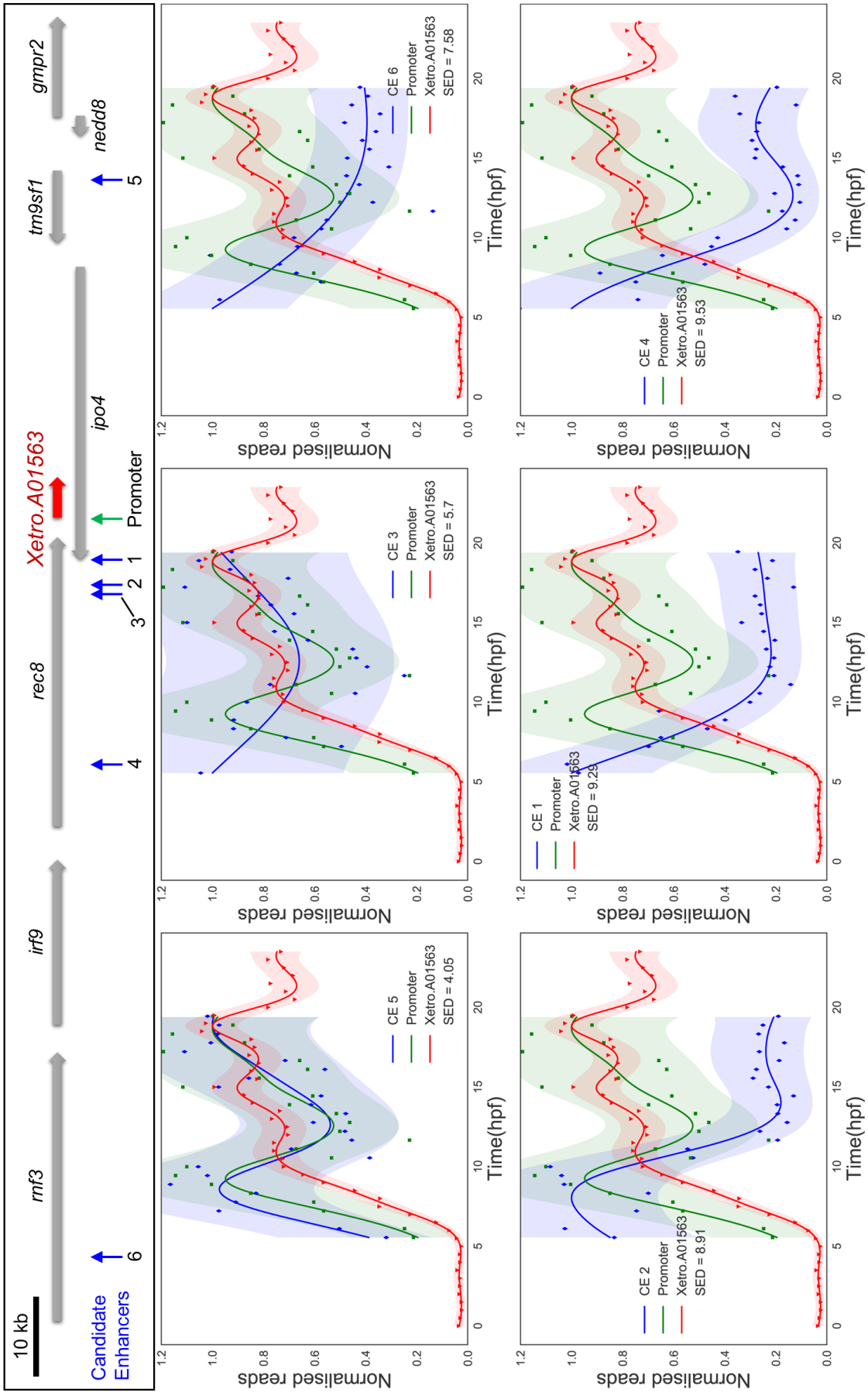

10 kb
Candidate Enhancers
6
rnf3
irf9
rec8
4
Xetro.A01563
ipo4
3
2
1
Promoter
tm9sf1
nedd8
5
gmpr2
CE 5
Promoter
Xetro.A01563
SED = 4.05
CE 3
Promoter
Xetro.A01563
SED = 5.7
CE 6
Promoter
Xetro.A01563
SED = 7.58
CE 2
Promoter
Xetro.A01563
SED = 8.91
CE 1
Promoter
Xetro.A01563
SED = 9.29
CE 4
Promoter
Xetro.A01563
SED = 9.53
Normalised reads
Time(hpf)

Figure 33 – Candidate enhancer-gene prediction over different genomic distances.

Example of a distant p300 region correlating better with a gene transcription that nearby ones. Schematic representation of the locus, with the several genes and candidate enhancers. Graphs with *Xetro.A01563*'s expression, p300 binding in its proximal promoter and in six candidate enhancers (CE), with the SED score for each trio. Graphs ordered by increasing SED. Candidate enhancers ordered by increasing distance to the *Xetro.A01563* proximal promoter.

4.6.1 Method testing

Nakamura and colleagues identified Wnt8a enhancers (enhancers bound by Wnt8a) and their target genes, in gastrula *X. tropicalis* embryos, by a combination of β-catenin ChIP-seq and Wnt8a morpholino knockdown, followed by RNA-seq (Nakamura et al., 2016). They tested the ability of seven of these predicted enhancer regions to drive gene expression by luciferase assay; four of the regions were positive and three were negative. This provided a small set of regions to validate the corresponding candidate enhancer-gene prediction method described in the previous section.

Table 10 summarises the tested regions and the results. The target genes for these candidate enhancers were *cdx2*, *cdx4*, *hoxa1*, *hoxd1* and *msx1*. *Hoxa1* (region 4) and *hoxd1* (regions 8 and 9) do not have detectable p300 binding at their proximal promoter regions, therefore they could not be analysed with the SED method, which requires p300 binding at the proximal promoter. Two of the seven tested regions were composed of two β-catenin ChIP-seq peaks (2 + 3 and 5 + 6); region 6 did not have a corresponding p300 region, therefore it was not analysed. After excluding those regions, the SED was calculated for regions 1, 2, 3, 5 and 7.

	β-catenin regions						p300 regions			
Region Number	**scaffold**	**start**	**end**	**gene**	**gene distance**	**Luciferase Assay**	**scaffold**	**start**	**end**	**SED**
1	scaffold 1	193640871	193641732	msx1	-2952	Positive	scaffold 1	193640895	193641456	5.48
2	scaffold 2	106293934	106294533	cdx2	-2470	Negative	scaffold 2	106294022	106294549	8.78
3	scaffold 2	106294856	106295371	cdx2	-1592		scaffold 2	106294863	106295376	8.94
4	scaffold 6	103541474	103542376	hoxa1	2956	Negative	no p300 at proximal promoter			
5	scaffold 8	26301044	26301471	cdx4	2923	Positive	scaffold 8	26301034	26301392	4.28
6	scaffold 8	26301633	26301948	cdx4	3435		no corresponding p300 region			
7	scaffold 8	26302751	26303306	cdx4	4616	Negative	scaffold 8	26302767	26303243	5.79
8	scaffold 9	20751050	20751532	hoxd1	6664	Positive	no p300 at proximal promoter			
9	scaffold 9	20754178	20755413	hoxd1	3295	Positive				

Table 10 – Regions for candidate enhancer-gene pairing method testing.

Nine β-catenin regions from Nakamura et al., 2016 tested as Wnt8a target enhancers, their genomic location, distance to target gene and luciferase assay result. These regions were searched for corresponding p300 regions and the SED score between the p300 binding dynamics at those enhancers and the target promoter and the corresponding gene's expression was calculated. Regions without p300 binding or that the corresponding promoter does not have p300 binding were excluded from the analysis.

Regions 1 and 5, which tested positive in the luciferase assay, had the lowest SED of the five regions (5.48 and 4.28 respectively, compared to 8.79, 8.94 and 5.79 for 2, 3 and 7, respectively, which tested negative).

Figure 34 shows the p300 binding profile at each of the five regions as well as the p300 binding at the target proximal promoter and the target gene's RNA abundance. The three profiles (gene, proximal promoter and candidate enhancer) are very similar for 1 and 5 (positive regions), as well as for 7, which tested negative in the luciferase assay, however it had a relatively low SED (high correlation). For both regions with high SED (2 and 3 – negative regions), the p300 binding dynamics at the candidate enhancer region are actually similar to the gene's expression dynamics. What increases the SED is the p300 binding at the proximal promoter, which is not similar to either the p300 at candidate enhancer or to the gene's expression. Given that these regions tested negative in the luciferase assay, it reflects the importance of including the dynamics in enhancers, promoters and genes in their method, to exclude candidates in cases where one of the dynamics does not match. At least in these examples, if the proximal promoter dynamics were not included, these candidate enhancers would correlate well with the gene, however it was experimentally demonstrated that these are likely not enhancer regions.

Even though five regions are a very small sample, the method yields lower scores for the two positive regions and higher scores for the three negative regions which is a positive indication that this method has good potential. It would be interesting to perform further testing to validate the method.

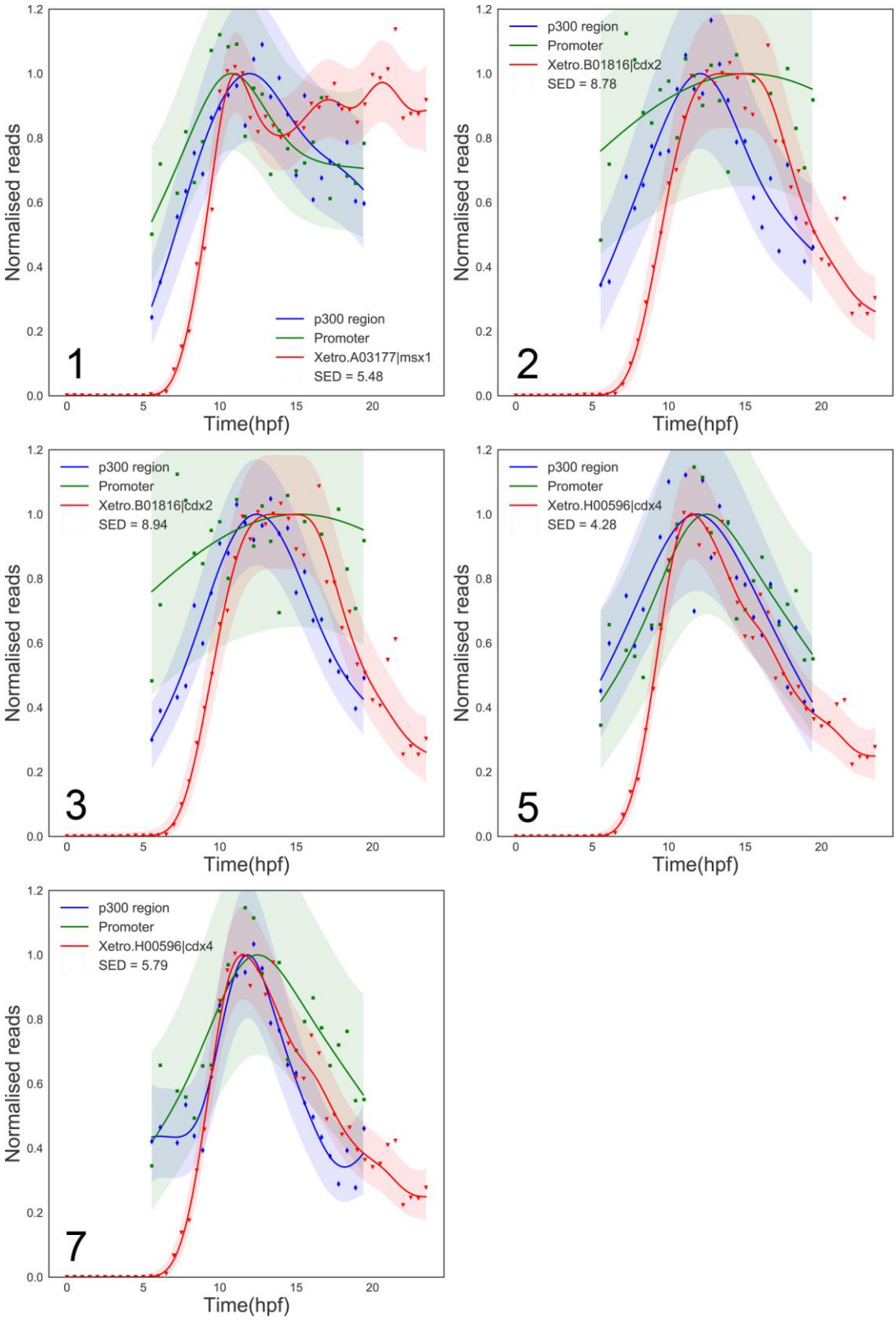


Figure 34 – Candidate enhancer-gene method testing.

p300 binding dynamics (blue) at the five tested enhancers (from Nakamura et al., 2016) and the binding in the corresponding proximal promoter (green) and target gene's expression (red).

4.7 p300 and eRNA dynamics

Enhancer RNAs were first discovered in 2010 (Kim et al., 2010) and since then several studies have showed that they are involved in transcriptional regulation (Mousavi et al., 2013, Iott et al., 2014, Schaukowitch et al., 2014). In this section I took advantage of the p300 and mRNA temporal data to evaluate eRNAs at the list of candidate enhancer regions generated in Chapter 3.

A set of candidate enhancers was selected, by taking all intergenic p300 regions (9,149 regions) and excluding those with H3K4me3 (based on the Hontelez et al., 2015 data), a promoter mark, to minimise the risk of including regions in unannotated promoters, leaving 5,847 regions (64%). The number of RNA-seq reads from the raw data from Owens et al., 2016, both from the ribosome-depleted and poly(A)-selected RNA-seq data, were counted and the previously described Gaussian process was applied to each region. To allow for detection of eRNAs independent of their polyadenylation status I focused on the ribosome-depleted data. Results were similar for the poly(A)-selected data (data not shown).

Regions that had less than, on average, 1 read per time point were excluded, with 1625 regions remaining (18% of the p300 regions). Kim and colleagues reported that 25% of candidate enhancers are transcribed into eRNAs, therefore, the resulting number of regions from this project's data is close to the expected. Hereafter, I will refer to RNAs expressed from candidate enhancer regions as eRNAs, however, some may actually represent unannotated gene regions.

The ED between the eRNA expression profile and the p300 dynamics for each region was calculated, as well as between eRNA and a random p300 region.

The ED between eRNAs and nearby genes (between 1 and 20 kb away – ensuring that eRNAs did not overlap annotated gene regions) was also calculated.

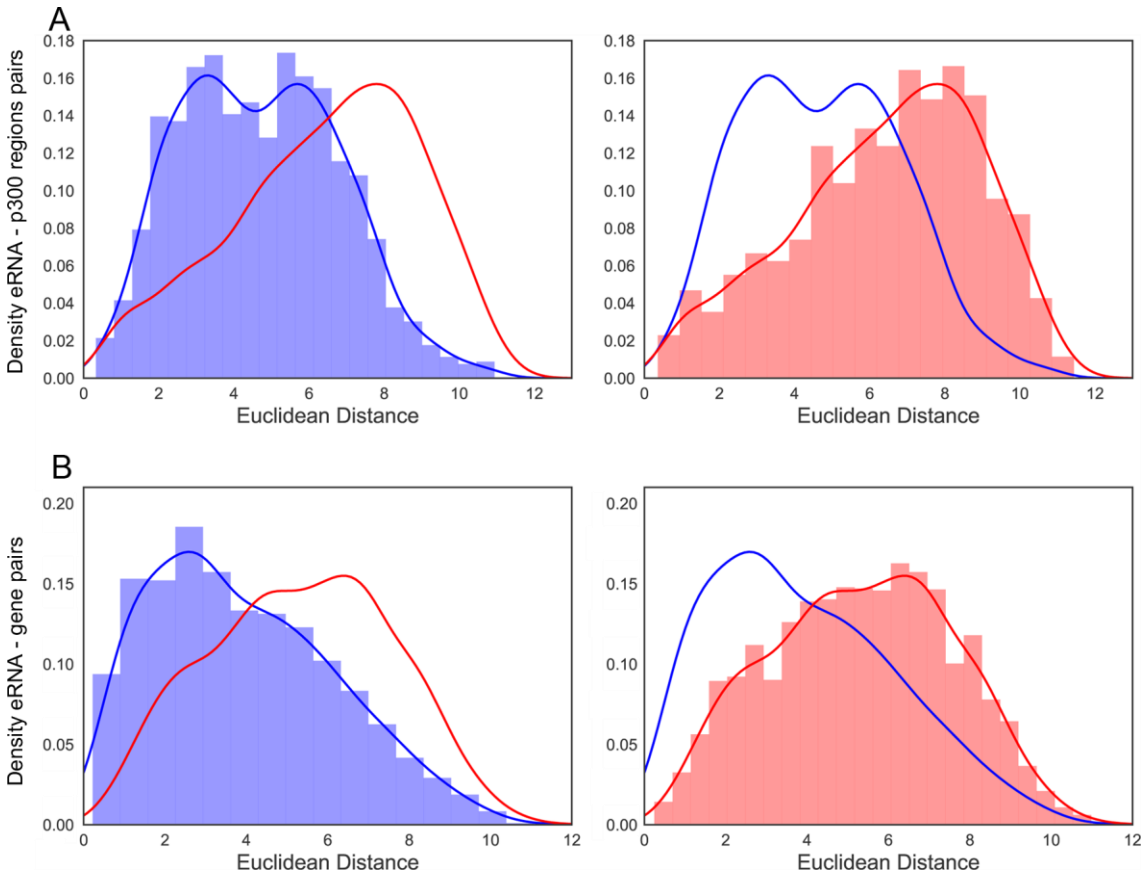


Figure 35 – Euclidean distances between eRNA expression and p300 binding/gene expression.

Histograms and distribution curves representing the EDs between ribosome-depleted RNA-seq reads in intergenic non-H3K4me3 p300 regions (eRNAs) and A – p300 binding at those same regions, B – Transcription levels for nearby genes. In red are the same comparisons but with random p300 regions/genes from a different chromosome.

eRNAs expression dynamics correlate well with both p300 binding dynamics in the same region (mean ED = 4.72) and with nearby gene expression (mean ED = 3.92) and, for both comparisons, it correlated better than with randomly assigned candidate enhancers/genes (27% decrease in mean ED for both, p-value = 6.46e-95 and 9.37e-82, respectively, Mann-Whitney U test). Figure

36 shows examples of four candidate enhancer regions, with the p300 binding and eRNA expression dynamics in each region. Regions 1 and 2 are 19 and 43 kb upstream of the nearest gene, respectively. Regions 3 and 4 are 267 and 25 kb downstream of the nearest gene, respectively.

It is interesting to note that the eRNA in region 4 is present from fertilisation. This could indicate that there are maternally deposited eRNAs, or that these are either inside unannotated genes or that the nearest gene's model is incorrect. The nearest gene to this region is *prr16* (25kb upstream of the eRNA), which is not present at fertilisation, therefore, at least in this case, the latter is not a likely option.

Overall, 9.5% (154 out of 1625) of eRNAs have their maximum at fertilisation. To my knowledge, maternally deposited eRNAs have never been reported, and it would be interesting to verify if these RNAs are indeed eRNAs.

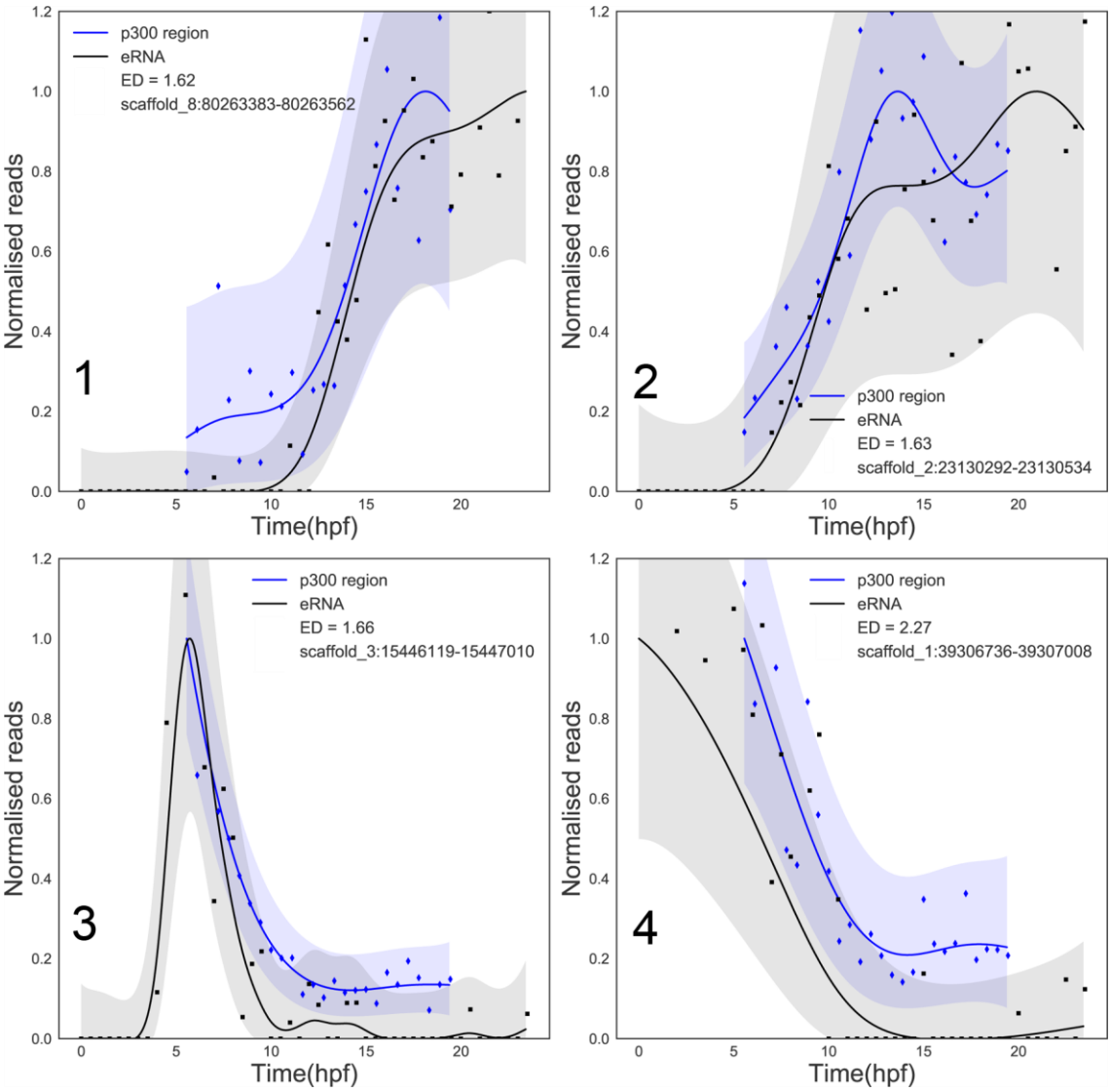


Figure 36 – eRNA and p300 binding dynamics.

Examples of p300 binding dynamics (blue) and eRNA expression (black) in four candidate enhancer regions, as well as each pair's ED.

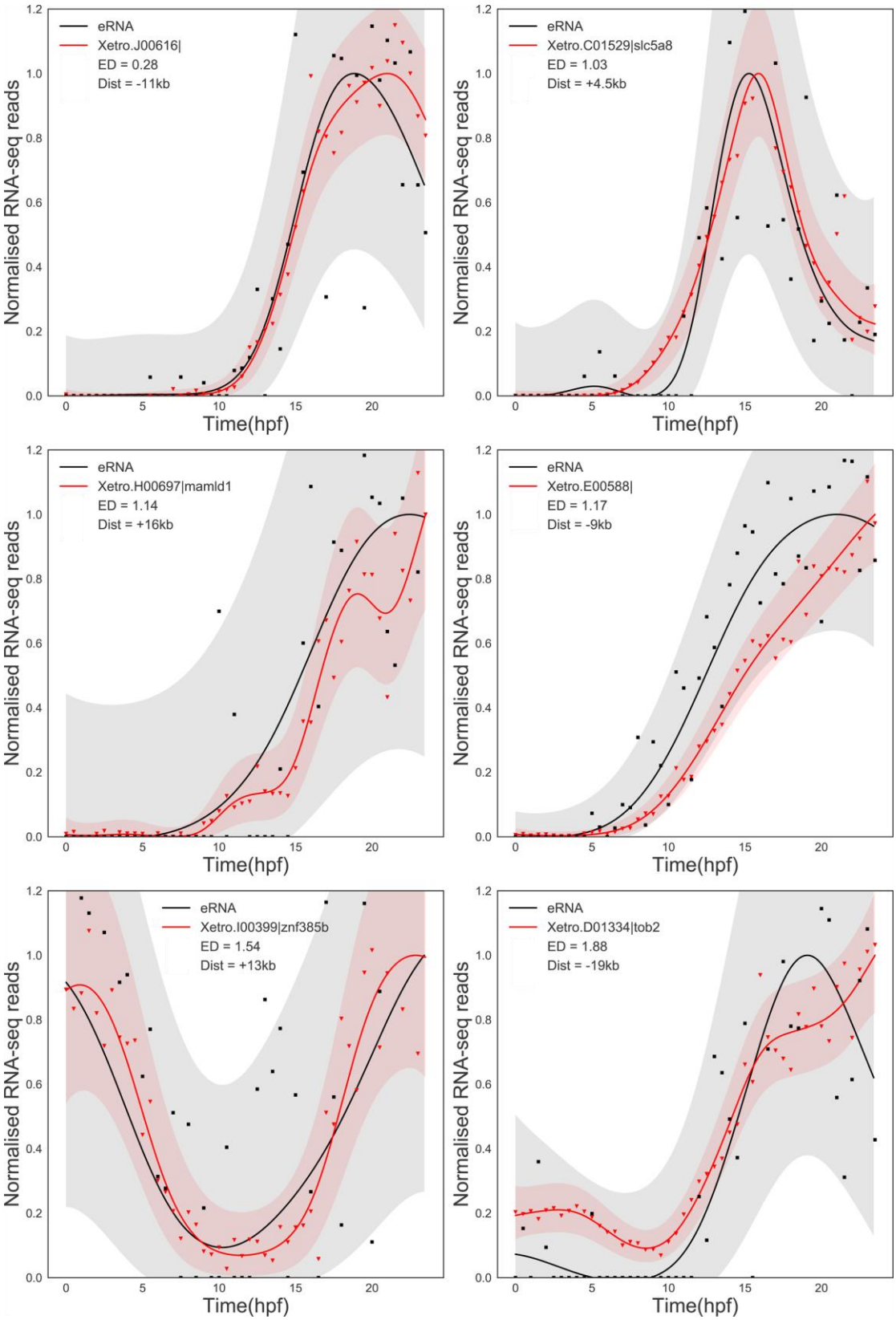


Figure 37 – eRNA and nearby gene expression.

Examples of eRNAs (black) and nearby genes' expression (red), as well as each pair's ED and the genomic distance between the two regions.

Figure 37 shows examples of eRNAs and nearby genes' expression. On average, eRNA expression correlates better with gene expression than with p300 binding (17% decrease in mean ED, p-value = 6.34e-26, Mann-Whitney U test), which would make eRNA dynamics a very interesting method to explore enhancer-gene pairs. However, this analysis is beyond the scope of this thesis. There are several factors confounding that complicate this eRNA analysis, including unannotated genes and other non-coding RNAs, poor gene models or alternate promoters.

4.8 Discussion

In this chapter I have described the relationship between p300 binding and gene transcription. I showed that the more transcriptionally active a gene is (higher expression level fold change), the more likely it is to have p300 binding nearby and that the binding itself is higher (more normalised ChIP-seq reads), showing a correlation between p300 binding and transcriptional activity.

On average, the net transcription rates of active genes match the timing of nearby p300 dynamics, with genes near early p300 regions being active at early time points, near mid p300 regions at the middle of the time series and near late p300 regions at late time points. This shows that, on a genome-wide average, a gene's transcription correlates with nearby p300 binding dynamics.

I then set out to determine if p300 binding dynamics at promoters and nearby (<20 kb) candidate enhancers correlate. Not all possible enhancer-promoter combinations within a 20 kb region will be part of a DNA loop interaction. The hypothesis was that it would correlate better than random pairings, as 20 kb interactions will be considerably more likely than interchromosomal ones. Therefore, it was expected that several nearby candidate enhancer-proximal promoter pairs would have low correlation scores (high EDs), however that on average they would be more correlated than random pairs. The analysis showed that, indeed, p300 binding dynamics at proximal promoters and at nearby candidate enhancers correlate well, and significantly better than if the proximal promoters are randomly paired with candidate enhancers in a different chromosome. This supports the model of p300 being involved in enhancer-promoter looping, due to the binding dynamics being similar on these paired regions.

Multiple candidate enhancers in the same loci may have the same p300 binding dynamics. This may represent a complex DNA-loop structure, with multiple enhancers interacting with the same promoter simultaneously, or each enhancer may be responsible for regulating transcription in different tissues. My data cannot answer this, but it would be interesting to perform p300 ChIP-seq on different tissues to clarify which of the two are responsible for the overall result.

I then set out to determine if p300 binding dynamics at proximal promoters also correlated well with the transcription of the corresponding gene. This would only be expected if looping is maintained during transcription, otherwise p300 binding would only correlate with the timing of transcription activation. p300 binding dynamics at proximal promoters does correlate with transcript levels, however not with the gene's net transcription rate. I had hypothesised that the more cells have p300 binding at a given promoter, the more that gene is transcribed (i.e. higher net transcription rate). p300 dynamics may not be an instantaneous indicator of transcription rate, due to the time required for, for example, RNAPII recruitment and release into the gene body. Another possible confounding factor is that net transcription rate describes active transcription rate *minus* mRNA degradation rate, however the ratio between the two is unknown and likely to vary between different genes and over time, as a gene is activated and builds up transcripts. Other methods which can be used to estimate transcription rates include measuring intronic reads in RNA-seq data (Madsen et al., 2015); RNAPII ChIP-seq time series (Maina et al., 2014); sequencing nascent RNA (GRO-seq) (Core et al., 2008, Jonkers et al., 2014); and sequencing chemically labelled RNA, for example with bromouridine (Veloso et al., 2014, Roberts et al., 2015).

Predicting which candidate enhancer regulates which gene's expression is not a straightforward task. This can be done experimentally for a small number of pairs, for example by 3C or 4C, and then validating the predicted regions by deleting them, for example by CRISPR-Cas9, however, it is not feasible to attempt it genome-wide. Making use of the high-resolution p300 binding data and the previously published high-resolution absolute RNA-seq data (Owens et al., 2016), I attempted to computationally assign candidate enhancers to their corresponding genes. Having shown that p300 binding at proximal promoters and at nearby candidate enhancers and gene transcription correlate well, and importantly, better than randomly assigned pairs, I used this information to predict which candidate enhancer regulates which gene.

The developed method – Sum of Euclidean Distances (SED) – adds the proximal promoter-candidate enhancer ED and the proximal promoter-gene ED, to generate a score for each candidate enhancer-gene pair within 100 kb of each other (the lower the score, the more correlated the pair is).

This method can be applied to very large distances, to discover candidate enhancers with a high correlation with a specific gene. The downside of scanning large regions is that the number of false positives will likely increase. This tool can be used by investigators to filter which enhancers to test experimentally.

According to this candidate enhancer-gene pairing prediction method, the closest candidate enhancer regulates only 25% of genes. Notably, this is in contrast to most annotation software that attributes candidate enhancer regions to the closest gene.

This method was tested against a small set of published enhancer-gene pairs, with 100% accuracy. The set size (five pairs) was very small and further

testing is necessary, however hardly any enhancer targets have been validated in *X. tropicalis* so far.

The SED method has some limitations, namely:

1. In cases where more than one enhancer regulates one promoter at different time points, it would not be expected that their binding dynamics would match perfectly, due to each enhancer contributing to part of the p300 promoter binding. Each enhancer would seem to partially correlate with the promoter, however none would have a perfect score. This is consistent with the finding from the previous chapter that p300 binding at promoters is less dynamic than at candidate enhancers. If the p300 binding at a promoter is approximately the sum of the p300 binding at the different enhancers, the promoter would be bound for longer periods. This could be tackled by calculating the ED for shorter time intervals instead of calculating the score for the whole time series, or by searching for trios of enhancer-enhancer-promoter, where the sum of the two enhancers correlates with the promoter.
2. Part of the random pairings are also well correlated (low ED), both when comparing promoter-candidate enhancer and promoter-gene. During early development, several genes and enhancers are activated by the same developmental processes, therefore it would be expected to find random pairs that correlate well, simply because the same cellular process is activating both at the same time. This will lead to some false positives, however this analyses showed that,

in general, the randomly generated pairs have significantly worse scores.

3. A candidate enhancer with extremely low ED (well correlated) with a proximal promoter could potentially be predicted as regulating a gene even if the proximal promoter-gene ED was very high (not correlated), due to the former being so low that the sum would be a relatively low value. This issue could be tackled by setting a minimum threshold for both EDs (candidate enhancer-proximal promoter and proximal promoter-gene) in order to predict an enhancer-gene pair.
4. Finally, only enhancer-gene pairs regulated by p300-mediated looping will be predicted with this method. As mentioned in the introduction, not all genes require p300 to be transcribed, therefore it would not be expected that all genes will correlate well with the p300 at its promoter.

I have also shown that RNA-seq reads from candidate enhancer regions, potential eRNAs, correlate well with both p300 binding in the same location and with nearby gene expression. If these RNAs are indeed eRNAs and not unannotated genes or due to incorrect gene models, this would be an interesting area to study further in an attempt to predict enhancer-gene pairs, by finding pairs with high eRNA-RNA correlation.

To my knowledge, this is the first time potential eRNAs are reported in *Xenopus*. Almost 10% of these RNAs are present from fertilisation, suggesting that they may be maternally deposited. As previously mentioned, there are several confounding factors for this analysis, which could be resolved in future studies.

In summary, I showed that p300 binding at candidate enhancers and proximal promoters correlates well, in addition to p300 binding at proximal promoters and that gene's expression. Using this information, I developed a candidate enhancer-gene pairing method and generated a list of most likely pairs for genes active during the times assayed in this work (7.1 – Appendix 1 – Predicted enhancer-gene pairs).

Chapter 5. p300 in *Xenopus* development

5.1 Introduction

The aim of the analysis described in this chapter was to investigate p300 binding in the context of *X. tropicalis* embryo development, to better understand the processes in which p300 is involved and to explore whether the dynamic data generated in this project can help predict candidate target genes for different transcription factors. I performed differential motif binding analysis, to determine which motifs are enriched in p300 regions active at different developmental stages.

Xenopus development has been extensively reviewed (for example Hausen and Riebesell, 1991, Nieuwkoop and Faber, 1994, Wolpert, 2011). Very briefly, *Xenopus* development starts with external fertilisation followed by 12 rapid and synchronous cell divisions, with low levels of zygotic transcription. After the 12th cell division (between stage 8 and 9), the Mid-Blastula Transition (MBT) occurs: the cell cycle desynchronises and lengthens, and cells become motile. The blastula stages comprise the period before the MBT, as well as a short time period post-MBT prior to gastrulation. At around stage 10, gastrulation starts and cells reorganise to establish the animal body plan, giving rise to the three germ layers:

1. Ectoderm – Precursor of the nervous system and epidermis
2. Mesoderm – Precursor of tissues such as bone, gonads, kidney and muscle and the circulatory system
3. Endoderm – Precursor of organs such as the digestive and respiratory systems organs

At around stage 13 neurulation starts. Several tissues develop and differentiate during these stages, the most prominent of these developmental

processes being neurogenesis, ending with the closure of the neural tube at stage 20.

5.2 p300 differential motif binding

p300 interacts with multiple transcription factors at enhancers, therefore motif analysis on p300 binding regions should yield a variety of motifs. I hypothesised that at early time points p300 should associate with motifs for maternally deposited transcription factors and later in development with motifs for zygotically transcribed transcription factors.

I performed differential motif analysis, to determine which transcription factor motifs are more associated with p300 occupancy at different developmental stages, using Gimmemotifs maelstrom (van Heeringen and Veenstra, 2011). This program combines the outcome of different statistical tests, regression and classification methods to calculate the association between transcription factor motifs and p300 occupancy in each time point and the resulting values represent the $-\log_{10}$(p-value) of the rank aggregation. A high value means that the motif is predicted to be positively associated with p300 binding for that time point. For example, Figure 38 shows high values for the Foxh motif in the first hour and a half of the time series, this means that p300 regions with this motif are more likely to have a high p300 read count in early time points and low read count in late time points.

Motifs for transcription factors not identified in *Xenopus* were excluded and the 10 transcription factor motifs with the highest differential association with p300 were selected for further analysis (Table 11 and Figure 38).

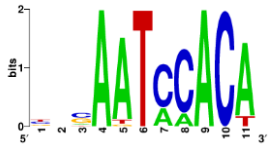
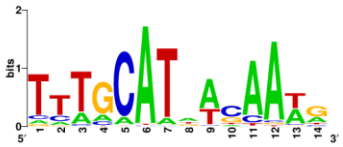
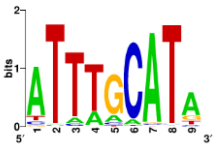
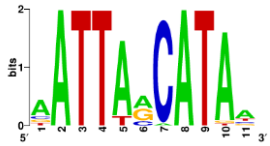
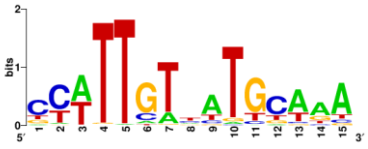
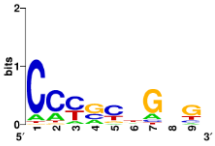
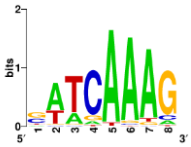
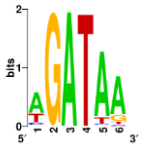
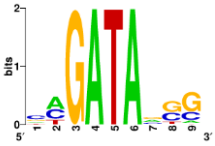
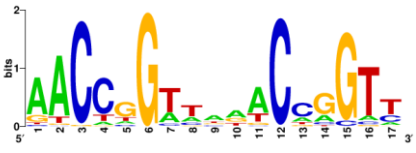
	Motif Name	Transcription factor family	Motif
1	Forkhead_M2151_1.01	Foxh	
2	Homeodomain_POU Average_47	Pou	
3	Homeodomain_POU Average_43	Pou	
4	Homeodomain_POU Average_32	Pou	
5	Sox_Average_66	Sox	
6	C2H2_ZF_Average_172	Zic	
7	Sox_Average_97	Tcf7/Lef	
8	GATA_Average_32	Gata	
9	Unknown_M3387_1.01	Lmo	
10	Grainyhead_M5316_1.01	Grhl	

Table 11 – Motifs with highest differential association with p300 binding.

Motif names, transcription factor family and motif logos from (Hontelez et al., 2015), based on (Weirauch et al., 2014).

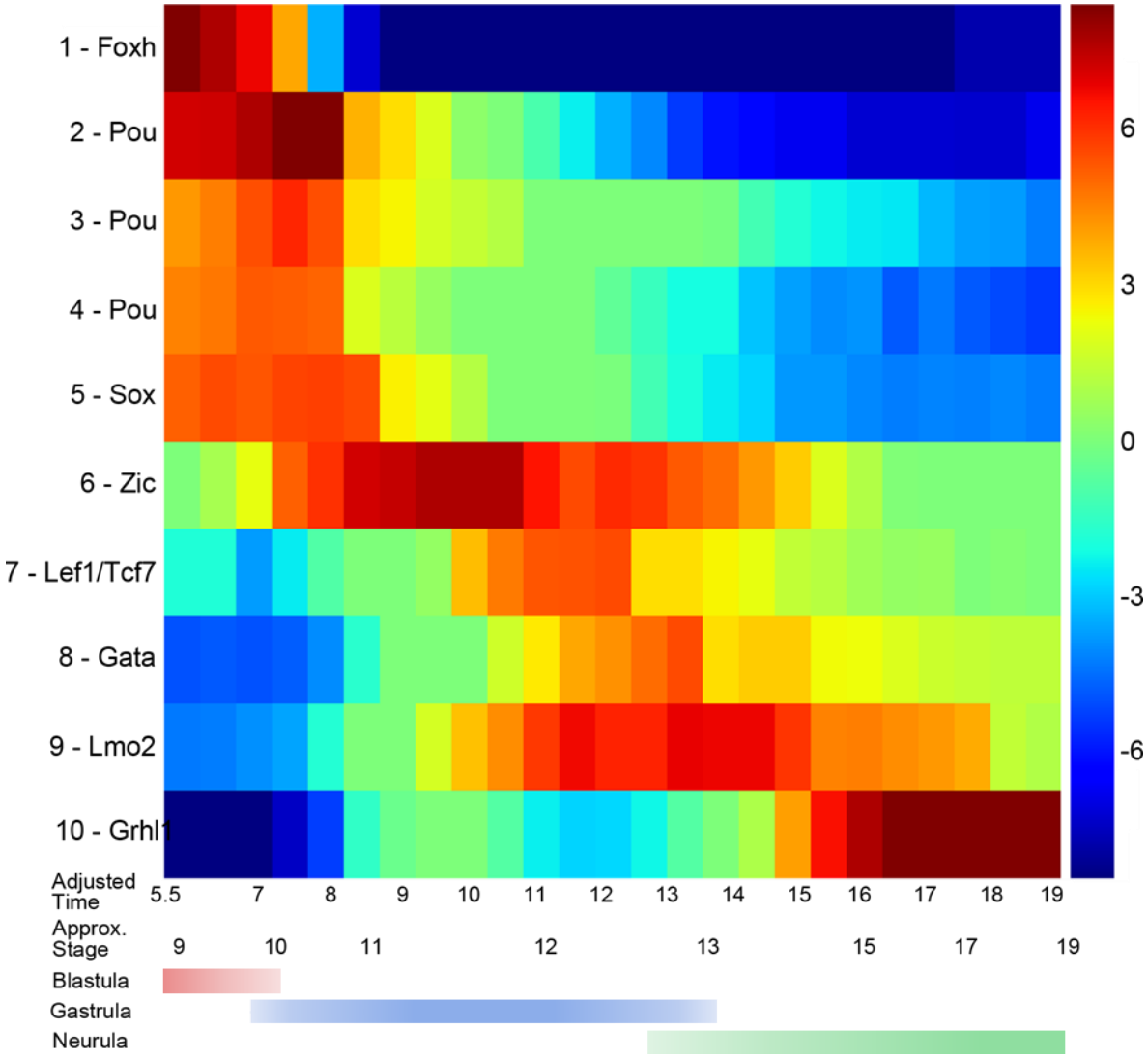


Figure 38 - p300 differential motif binding at different developmental stages.

A high value (red) indicates that regions with that motif have higher than average p300 reads at that time point.

The three Pou motifs are very similar, although motif 2, which has the strongest association with p300 of the three, is also the more degenerate. The Gata and Lmo motif are nearly identical (Table 11). These two sets of motifs were then analysed as a whole (the three Pou motifs as one and the Gata and Lmo as another).

In order to predict which transcription factors that bind to a given motif are more likely to be associated with p300, the time dependent behaviour of each motif was compared to the expression profile of the genes which encode the corresponding transcription factors, using the data from Owens et al., 2016.

A transcription factor can be present in the embryo either by being maternally deposited in the form of mRNA and/or protein or by being zygotically transcribed. The Owens et al., 2016 data describes the mRNAs that are present in the embryo, however, to determine if a transcription factor protein is maternally deposited it is necessary to analyse proteome data. There is no available data on the *X. tropicalis* proteome, so I searched the closely related *X. laevis* proteome to determine whether a protein was likely to be present in the egg. The catalogue by Wuhr and colleagues is by far the most extensive dataset available, they identified more than 11,000 proteins in the *X. laevis* egg (Wuhr et al., 2014). Other studies have much lower yields – 5,800 (Smits et al., 2014), 4,000 (Sun et al., 2014) and 1,400 (Sun et al., 2016) proteins. The absence of a protein in the Wuhr and colleagues data does not mean the protein is not present in the egg, only that it is under the detection limit.

5.2.1 Foxh motif

Foxh (forkhead box H, or Fast – forkhead activin signal transducer) motif is highly positively associated with p300 binding for the first 1.5 hours of the time series, which corresponds to the end of the blastula stages. *Foxh1* mRNA is maternally deposited and its zygotic transcription starts at around 3 hpf. Its expression starts decreasing sharply at around 5 hpf, with a second activation event at around 9 hpf. *Foxh1.2* mRNA was not detected in the oocyte and starts being transcribed at 4 hpf, however its expression levels are almost 20 times lower than *foxh1* (Figure 39). Foxh1 and Foxh1.2 were not detected in the *X.laevis* proteome data. Due to their expression profiles, Foxh1 is the most likely candidate to be associated with p300 at these early time points, given that *foxh1.2* starts being transcribed just before the start of the p300 time series. Foxh motif-p300 association has its maximum at the first time point of the time series, therefore they are likely associated for some time before that, even though the data cannot confirm that.

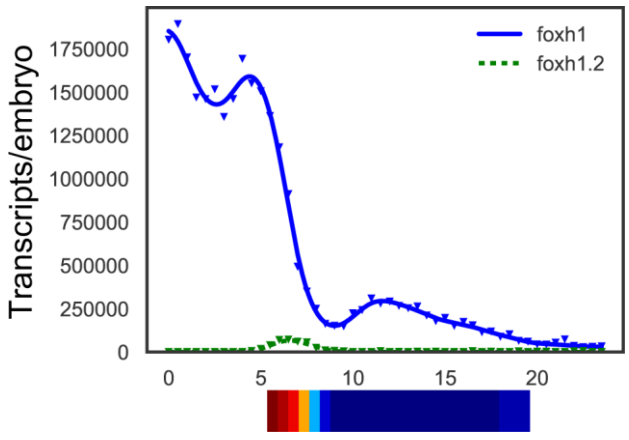


Figure 39 – *Foxh1* and *foxh1.2* RNA abundance and their motif association with p300 over time.

Dashed line indicates that that transcription factor is unlikely to be associated with p300, due to its RNA abundance over time not matching the p300 association timings. RNA abundance data from Owens et al., 2016.

Foxh1 is a member of the winged helix transcription factor family and it is necessary for the early stages of mesoderm specification and also to activate a number of endodermal genes (Chen et al., 1996, Watanabe and Whitman, 1999). Chen and colleagues identified Foxh1 as an interaction partner of Smads in *Xenopus*, being involved in TGFβ (Transforming growth factor β) signalling (Chen et al., 1996). Foxh1 is not able to drive transcription on its own (Chen et al., 1996), however, in the presence of TGFβ signalling, Smad2/Smad4 are translocated to the nucleus, binds Foxh1 and the complex is able to activate transcription (Chen et al., 1997, Liu et al., 1997). p300 is known to bind and acetylate Smads, which is required for Smad-mediated transcription (Ross et al., 2006, Inoue et al., 2007).

5.2.2 Pou motif

Three Pou motifs are positively associated with p300 binding during the first 2.5 hours of the p300 ChIP-seq time series, which corresponds to the end of the blastula stage and start of gastrulation. The three motifs are very similar, therefore they were analysed together.

There are 18 *pou* genes identified in *X. tropicalis*, however the only with maternally deposited mRNA are *pou2f1*, *pou5f3.2* and *pou5f3.3* (Figure 40). In the *X. laevis* proteome data, only Pou2f1 was detected at fertilization. *Pou5f3.1* is activated at 4 hpf, while the other members of the family are only expressed after around 10 hpf (Figure 40).

Pou5f3.2 and Pou5f3.3 are the most likely candidates to be associated with p300 binding at these stages, due to their expression profiles matching the period of high association with p300. However, Pou5f3.1 and Pou2f1 cannot be excluded given that they are present in the embryo at those times. Their expression peaks at later time points (11 and 18 hpf, respectively), however they may still associate with p300 at earlier times and lose their association due to other factors, such as a potential loss of a necessary cofactor at later stages.

Not much is known about Pou2f1 function in Xenopus development; there are indications that it may be involved in neural development (Veenstra et al., 1995) and specifically in radial glia formation (Kiyota et al., 2008), however these cells are only identified after stage 23 (Messenger and Warner, 1989) and this protein is maternally deposited, thus it may have a still unknown role in the early stages of development. Pou5f3.1, Pou5f3.2 and Pou5f3.3 are involved in maintaining cell pluripotency during gastrulation (Cao et al., 2004, Cao et al., 2006,

Morrison and Brickman, 2006, Nishitani et al., 2015). p300 is known to colocalise with Pou5f3.1 in enhancer regions (Chen et al., 2008).

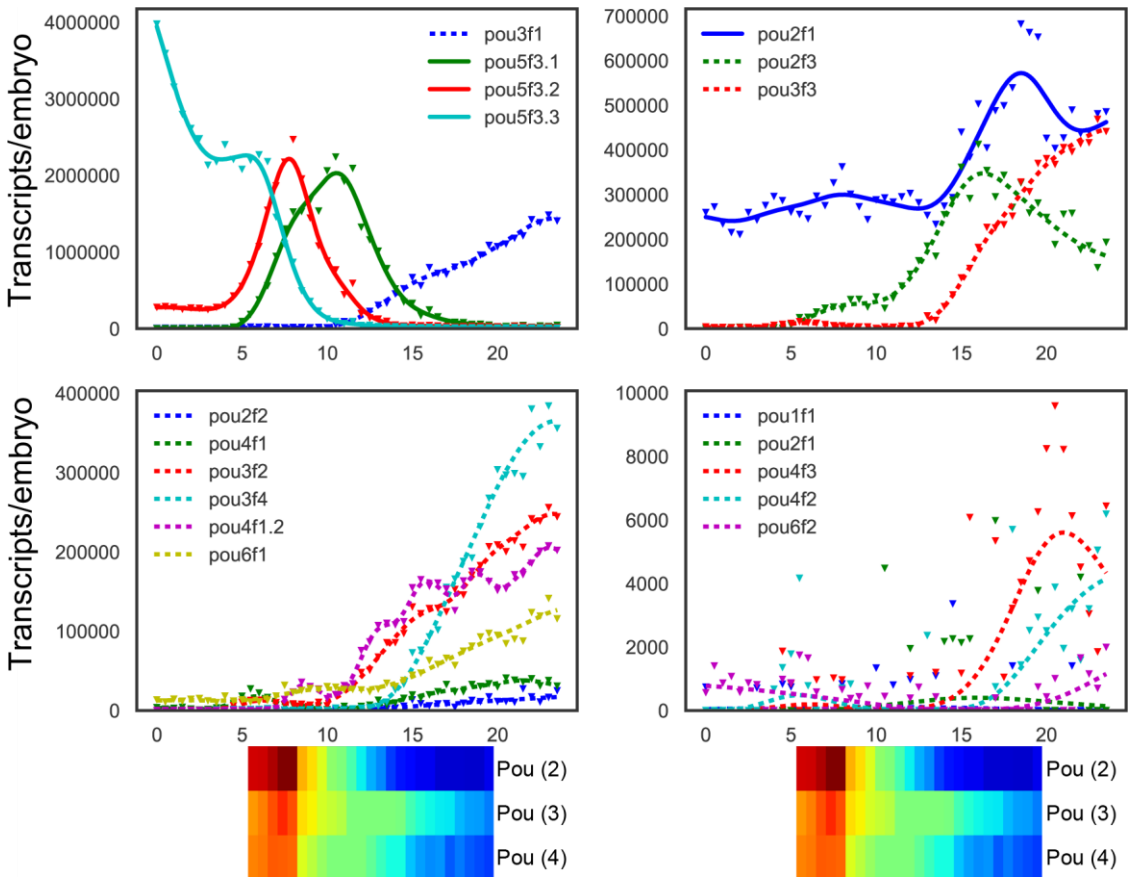


Figure 40 – *Pou* RNA abundance and their three motifs' association with p300 over time.

Dashed lines indicates that that transcription factor is unlikely to be associated with p300, due to its RNA abundance over time not matching the p300 association timings. Genes ordered by expression levels. RNA abundance data from Owens et al., 2016.

5.2.3 Sox motif

The Sox (SRY-box) motif is associated with p300 binding during the first 3 hours of the time series, corresponding to the end of the blastula stages and the start of gastrulation. From the Sox family of HMG (high mobility group)-box transcription factors, *sox3*, *sox7*, *sox11* and *sox13*, and low levels of *sox4* and *sox9*, mRNAs are maternally deposited, with some members of the family being transcriptionally activated before 5 hpf (Figure 41). Sox3 and Sox13 proteins are maternally deposited in *X. laevis* embryos.

For the reasons explained in 5.2.2 – Pou motif, the most likely candidates to be associated with p300 binding in regions with the Sox motif at these stages are Sox3, Sox7, Sox11 and Sox17b, however Sox2, Sox13 and Sox17a cannot be excluded.

Most Sox proteins are involved in neural development, however some members of the family are also involved in early embryo patterning and axis formation. Some examples are Sox3 and Sox7's functions in early development. Sox3, the gene with the highest maternal contribution, regulates dorsal axis formation (Zhang et al., 2003) and activates *sox2* transcription (Rogers et al., 2009). Sox7, also maternally deposited, is known to activate some of the earliest zygotic genes, such as the *nodals* and *mixer*, involved in mesoendoderm development (Zhang et al., 2005). At least Sox2 and Sox4 are known to interact with p300 to activate transcription (Nowling et al., 2003, Inoue et al., 2016).

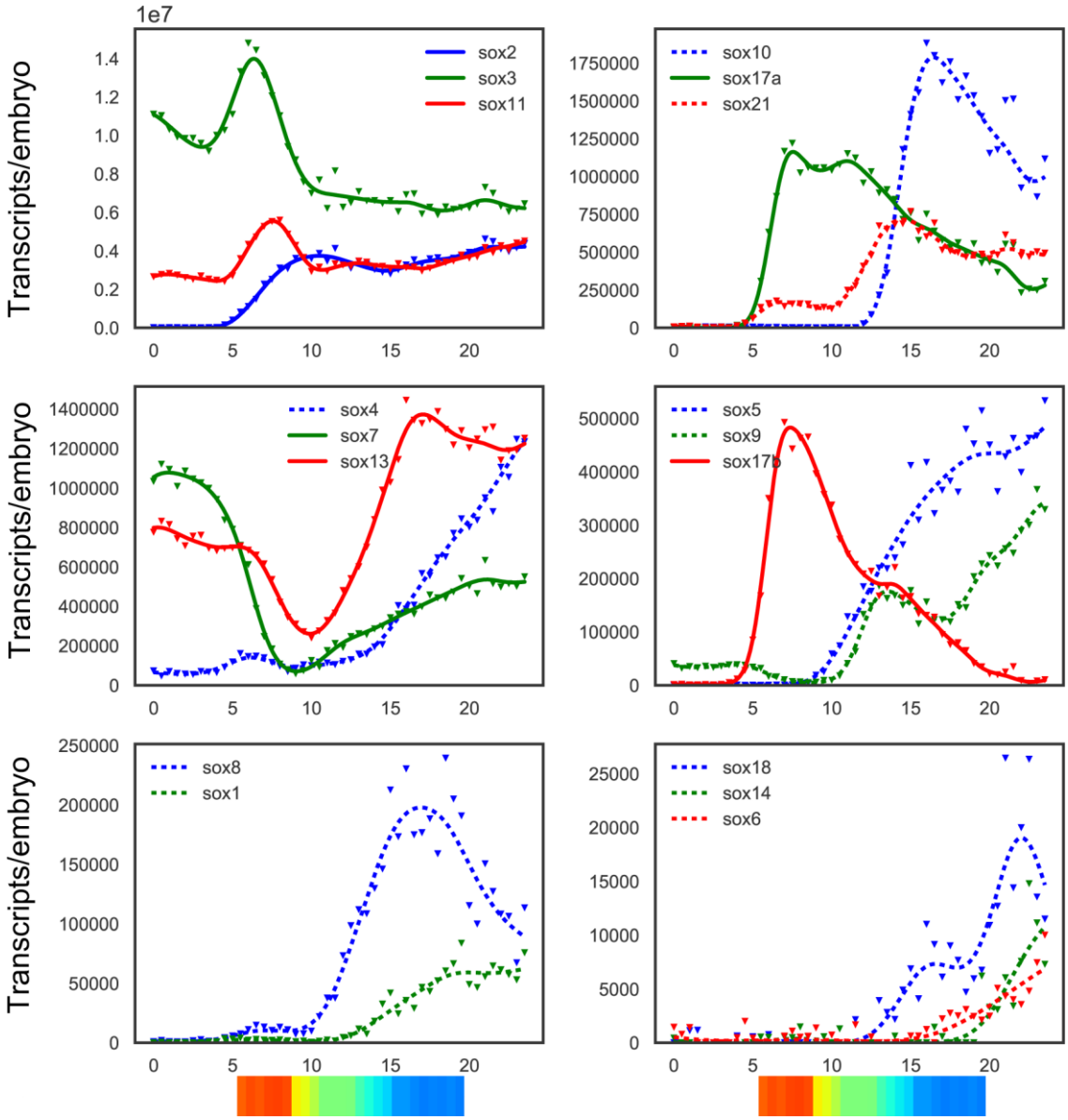


Figure 41 – *Sox* RNA abundance and their motif association with p300 over time.

Dashed lines indicates that that transcription factor is unlikely to be associated with p300, due to its RNA abundance over time not matching the p300 association timings. Genes ordered by expression levels. RNA abundance data from Owens et al., 2016.

5.2.4 Zic motif

The Zic motif is associated with p300 binding during gastrulation, from about 7 to 15 hpf, with a maximum association at around 10 hpf, which mainly corresponds to gastrulation. *Zic2* mRNA is maternally deposited (as well as low levels of *zic5*), while the other members of the family are activated at around 4 hpf. *Zic5* has a second wave of activation at about 8 hpf (Figure 42). No Zic proteins were found in the *X. laevis* egg.

Zic1 and Zic3's expression are the most consistent with the association with p300, making them the most likely candidates, however none of the other Zic proteins can be excluded.

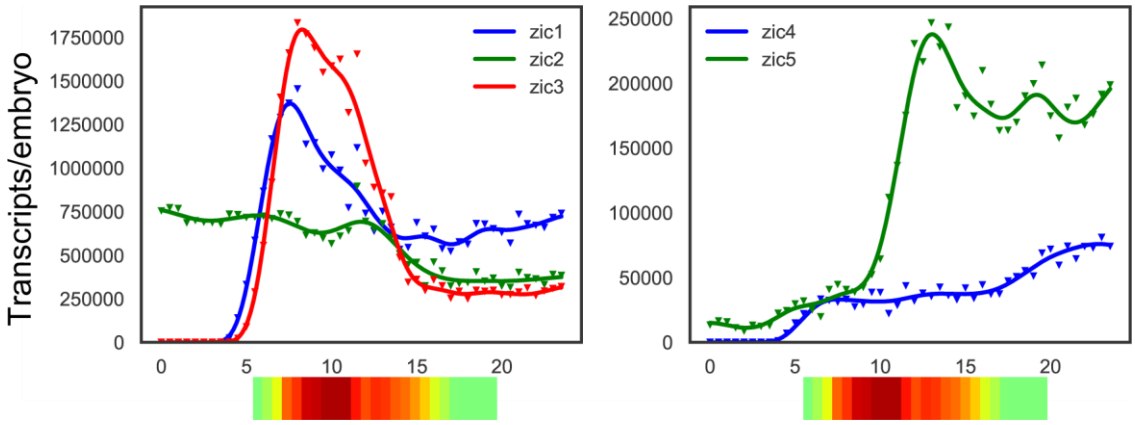


Figure 42 – *Zic* RNA abundance and their motif association with p300 over time. Genes ordered by RNA abundance. RNA abundance data from Owens et al., 2016.

These proteins are involved in inducing neuroectoderm and in regulating neural crest formation during neurulation (Nakata et al., 1997, Kuo et al., 1998, Nakata et al., 1998, Kitaguchi et al., 2000, Fujimi et al., 2006). As far as I am aware, there are no reported studies showing Zic-p300 interactions.

5.2.5 Tcf7/Lef1 motif

The Tcf7/Lef1 (T-cell factor/lymphoid enhancer factor) motif is positively associated with p300 binding for a short time window from 10 to 13 hpf, at the end of gastrulation.

Tcf7l1 and *tcf7* mRNAs are present in the oocyte, *lef1* is activated very early, at about 3 hpf, and *tcf7l2* at 13 hpf (Figure 43). Only Tcf7l1 protein was identified in the egg. Due to their presence in the embryo before 10 hpf, Tcf7, Tcf7l1 and Lef1 are the most likely members of the family to be interacting with p300 during this time interval, however none of their expression profiles peak during the high-association period. There may be necessary cofactors which are only present in this short time window that leads to the high p300 interaction, however my data cannot answer this.

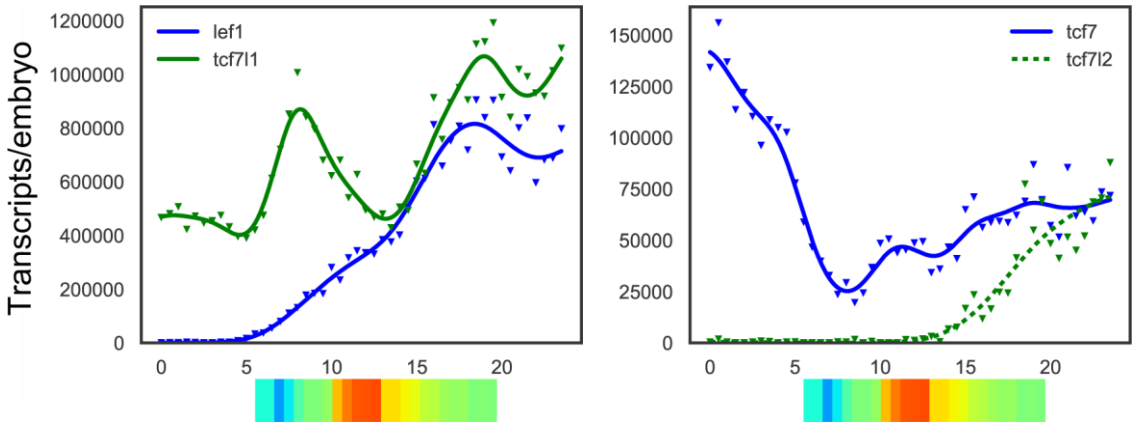


Figure 43 – *Tcf/lef* RNA abundance and their motif association with p300 over time.

Dashed line indicates that that transcription factor is unlikely to be associated with p300, due to its RNA abundance over time not matching the p300 association timings. Genes ordered by expression levels. RNA abundance data from Owens et al., 2016.

Tcf/Lef are HMG box transcription factors involved in Wnt signalling and can either activate or repress transcription. Lef1 is involved in mesoderm and ectoderm patterning during gastrulation (Roel et al., 2002, Roel et al., 2009); Tcf7 and Tcf7l1 are necessary for mesoderm induction (Liu et al., 2005) and the latter is also involved in the establishment of the dorsal axis (Roel et al., 2002). These transcription factors activate gene transcription by interacting with β-catenin, which in turn recruits co-activators, such as p300/CBP (Sun et al., 2000, Wolf et al., 2002, Ma et al., 2005), which has also been shown in *Xenopus* (Hecht et al., 2000, Takemaru and Moon, 2000).

5.2.6 Gata/Lmo motif

The Gata motif is associated with p300 binding during a short time window, between 11 and 13 hpf, corresponding to the end of gastrulation and beginning of neurulation. The Lmo (LIM domain only) motif is associated with p300 binding from 10 to 18 hpf, with the maximum association occurring between 11 and 15 hpf, corresponding to the end of gastrulation and neurulation. The two motifs are extremely similar, however their p300 association times have interesting differences. They both peak at around 13 hpf, however the Lmo motif association is present during a longer time interval.

Lmo4.1, *lmo4.2* and *lmo7* are maternally deposited, as well as g*ata5* and *gata*6, although at much lower levels. No Gata or Lmo protein was detected in the oocyte. From all the transcription factors in this families, only Gata1 can be excluded from interacting with p300 at these times, given that it is only activated at around 18 hpf, however the most likely candidates are Lmo2, Lmo4.1, Lmo4.2 and Gata2.

Lmo4 has been shown to be a cofactor for GATA proteins, involved in ventral mesoderm specification (de la Calle-Mustienes et al., 2003), which gives rise to, among other tissues, hematopoietic cells (Davidson and Zon, 2000).

Gata2 and Gata3 are involved in the segmentation of non-neural ectoderm (Read et al., 1998) and, as well as Lmo2, in early hematopoietic cells specification (Ting et al., 1996, Tsai and Orkin, 1997, Mead et al., 2001), which starts in the late gastrula/early nerula stages (Zon, 1995). Gata4, Gata5 and Gata6 are involved in heart (Jiang and Evans, 1996, Gove et al., 1997, Kuo et al., 1997, Nemer et al., 1999) and gut development (Arceci et al., 1993, Gao et al., 1998, Weber et al., 2000). Lmo7's function has not been studied in *Xenopus*.

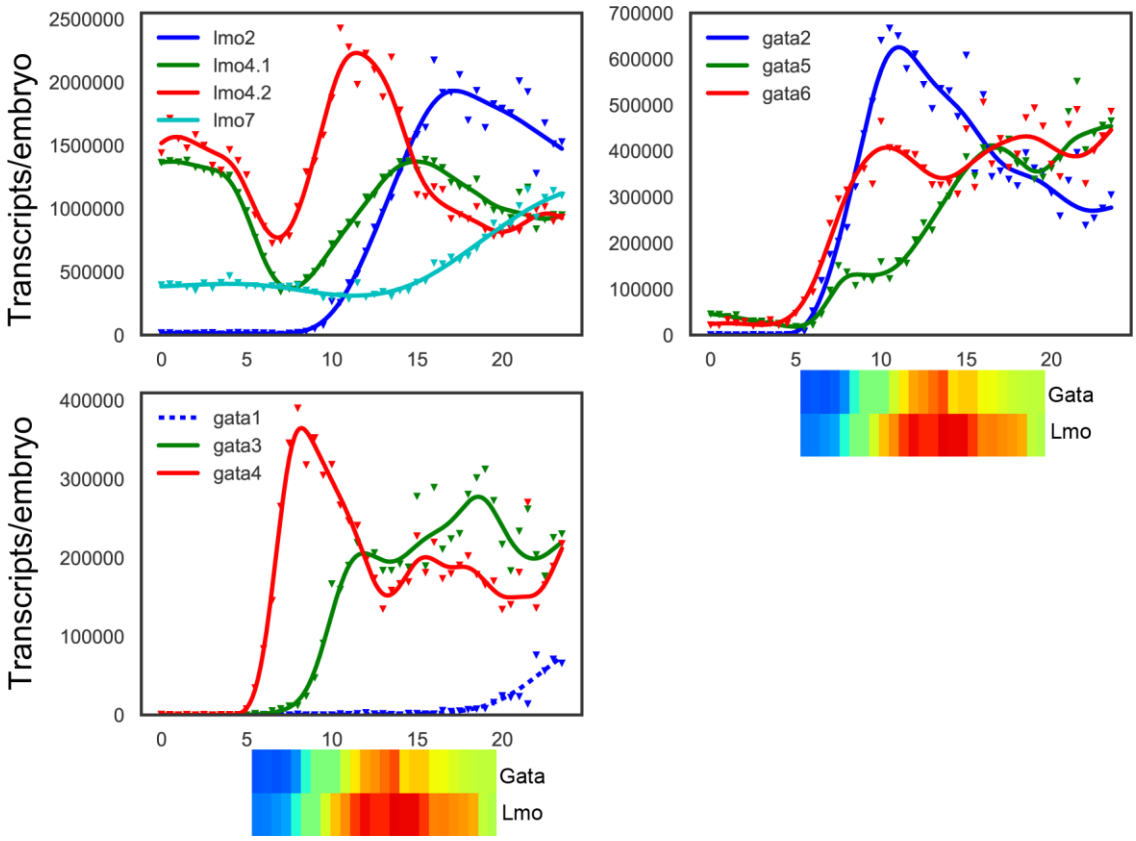


Figure 44 – *Gata* and *lmo* RNA abundance and their motif association with p300 over time.

Dashed line indicates that that transcription factor is unlikely to be associated with p300, due to its RNA abundance over time not matching the p300 association timings. Genes ordered by expression levels. RNA abundance data from Owens et al., 2016.

5.2.7 Grhl motif

The Grhl (Grainyhead-like) motif is highly associated with p300 binding in the final 3.5 hours of the time series, already during neurulation. *Grhl1* mRNA and protein is present in the fertilised oocyte, however its levels start decreasing immediately, with zygotic transcription of this gene starting at around 8 hpf. *Grhl2* and *grhl3* are activated at around 8 and 4 hpf, respectively (Figure 45). None of these genes show a peak in expression during the timing of high p300-association, however the three mRNAs are present and are potential candidates for p300 interaction.

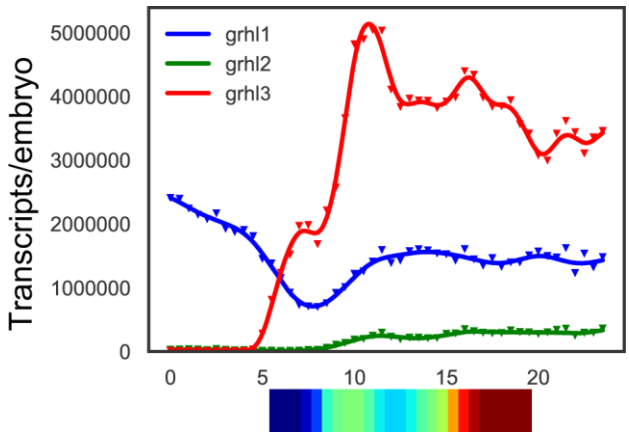


Figure 45 – *Grhl1* RNA abundance and its motif association with p300 over time. RNA abundance data from Owens et al., 2016.

Grhl1 and Grhl3 regulate epidermis differentiation as part of the BMP4 (Bone morphogenetic protein 4) signalling pathway (Tao et al., 2005), which is known to be involved in stimulating epidermis differentiation and inhibiting neural specification (Wilson and Hemmati-Brivanlou, 1995). Grhl2's function has not been assessed in *Xenopus*.

Suprisingly, all members of this family have been reported as p300-inhibitors, by inhibiting its HAT activity and p300's ability to interact with other

proteins (Pifer et al., 2016). Even though my data cannot shed light on this, it is possible that, given that these mRNAs are present at such high levels in early development, particularly *grhl1*, which is maternally deposited, their proteins may be inhibiting p300 from binding to these regions. An interesting possibility is that they may, then, not be bound at the later time points when p300 is recruited by another transcription factor present in the same enhancers. Further work would be required to evaluate this hypothesis.

5.2.8 p300 differential motif binding summary

p300 appears to bind different transcription factors at different developmental stages and those transcription factors are involved in very diverse developmental processes. For most of the motifs associated with p300 binding, the association timings are highly concordant with the transcription dynamics of certain members of the corresponding transcription factor family.

p300 appears likely to be involved in the early stages of mesoderm and endoderm specification, through its association with Foxh1 during the blastula stages; in pluripotency maintenance during gastrulation, by interaction with Pou proteins; in dorsal axis formation and activation of some of the earliest zygotic genes, such as *sox2*, *nodals* and *mixer* by interaction with Sox proteins; in neuroectoderm formation by interaction with Zic proteins; in germ layer induction and patterning in early *Xenopus* development, by interactions with β-catenin and Tcf/Lef proteins; and in hematopoietic cells differentiation and non-neural ectoderm patterning by interactions with Lmo and Gata proteins. Finally, p300 interacts with Grhl motifs during neurulation, however these protein have been reported as p300-inhibitors, by inhibiting its HAT activity and p300's ability to interact with other proteins (Pifer et al., 2016). The differential motif analysis suggests that p300 is enriched in regions with the Grhl motif, at a time when these protein are involved in epidermis differentiation, however p300's function may actually be inhibited in these regions. In light of this, it would be interesting to investigate if p300 is bound to these candidate enhancers at the same time and in the same tissues as the Grhl proteins.

Importantly, p300 appears to mark enhancers involved in all of these developmental processes.

5.3 Predicting transcription factor candidate target genes

I then sought to determine if the motif enriched candidate enhancers identified in this chapter can be combined with the enhancer-gene pairing method described in the previous chapter, to predict which transcription factors regulate which genes.

As an example, I focused on p300 regions containing the Foxh motif as, unlike many of the other motif families studied, Foxh1 appears to be the sole transcription factor responsible for the positive association of p300 and Foxh motifs at early time points. The Foxh motif is also one of the most highly associated with p300 and Foxh1 has been extensively studied, allowing the comparison of the results with previously published data.

The most representative regions of the p300-Foxh motif association were selected, by filtering the motif score from Gimmemotifs (hereafter called p300-Foxh1 candidate enhancers). Charney and colleagues performed Foxh1 ChIP-seq on early *X. tropicalis* embryos (Charney et al., 2017). 99% of the 1324 p300-Foxh1 candidate enhancer regions were detected by their study, showing that this method had a high specificity. Their study detected 40,884 Foxh1 regions, only a small fraction of that (3%) was detected with this method (low sensitivity). p300-Foxh candidate enhancers are 2.25 times more likely to have a Foxh1 ChIP-seq peak than non-p300-Foxh candidate enhancers (p-value = 0.0, Fisher Exact test).

The p300-Foxh1 candidate enhancers were analysed with the enhancer-gene prediction method described in the previous chapter, to determine which genes have the highest correlation (lowest SED) with those regions. The correlation between the p300-Foxh1 candidate enhancers and any gene within the

surrounding 1 Mb was calculated, given that there are examples of enhancers acting at those distances (Lettice et al., 2003). This distance can be adjusted to the characteristics of different experiments.

5.3.1 Candidate Foxh1 target genes

The top Foxh1 candidate target gene is *cdk9* (SED = 2.68), which, to my knowledge, has not been reported as a Foxh1 target. Both *foxh1* and *cdk9* are present in the ectoderm and endoderm in early gastrulation (Howell et al., 2002, Zhu et al., 2009), making this transcription factor-target gene pair plausible. The highly associated p300-Foxh1 candidate enhancer is located nearly 10 kb away from the *cdk9* promoter, in an intron of a nearby gene (Figure 46).

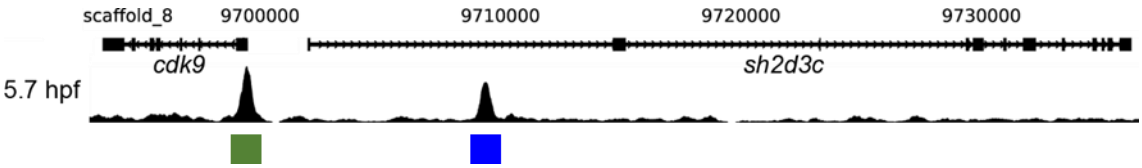


Figure 46 – *Cdk9* genomic locus.
Genome browser view of the *cdk9* and *sh2d3c* gene, with the two p300 regions predicted to be regulating *cdk9* through p300-Foxh1 interaction. *Cdk9* promoter region marked with a green box and p300-Foxh1 candidate enhancer with a blue box.

Figure 47A shows the *cdk9* expression profile (red) and the p300 binding dynamics at the gene promoter (green) and at the p300-Foxh1 candidate enhancer (blue). Figure 47B also shows the *cdk9* expression profile (red), as well as the *foxh1* gene expression profile (purple). All curves are normalised to their maxima.

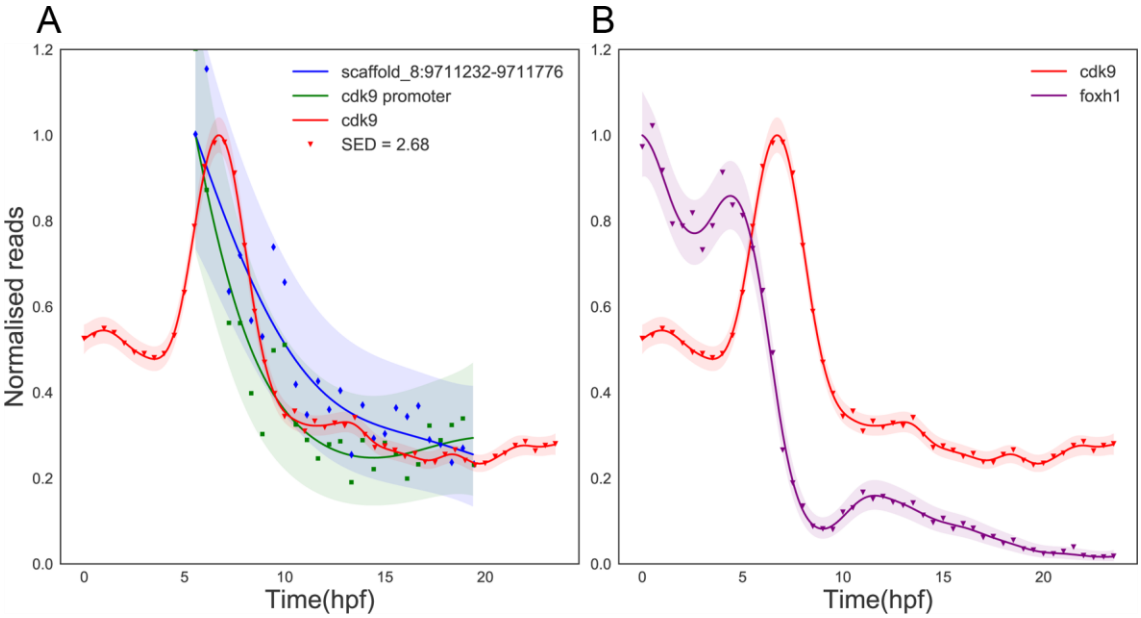


Figure 47 – *Foxh1* and candidate target gene *cdk9*.

A – *Cdk9* normalised expression profile (red) and normalised p300 binding dynamics in a p300-Foxh1 candidate enhancer (blue) and in the *cdk9* promoter (green). B - *Cdk9* (red) and *foxh1* (purple) normalised expression profiles.

Both the promoter and the p300-Foxh1 candidate enhancer were detected in a Foxh1 ChIP-seq experiment in early *X. tropicalis* embryos (Charney et al., 2017).

Chiu and colleagues performed a *foxh1* morpholino knockdown on early gastrula embryos (Chiu et al., 2014) (slightly later than the p300-Foxh1 association described in this chapter). *Cdk9* expression was not reduced by *foxh1* knockdown. From the top 10 candidate target genes, only two had reduced expression, *gdf3* (SED = 3.19) and *sds* (SED = 3.75) (for scale of SED data, see Figure 48).

To test the method with a known Foxh1 target, I analysed the results for *mix1. Mix1* is a known Foxh1 target (Chen et al., 1996, Chen et al., 1997) and its expression was reduced in the previously mentioned morpholino knockdown study (Chiu et al., 2014).

Both the *mix1* promoter and a candidate enhancer nearby contain the Foxh1 motif (and are present in the published Foxh1 ChIP-seq data). The SED for this candidate enhancer-*mix1* pair was low (5.95), being the 47th Foxh1 candidate target gene. Figure 48 shows the distribution of SED values for all p300-Foxh1 candidate enhancers, with the value for the *mix1* pair highlighted in red, showing it was one of the most correlated p300-Foxh1 candidate enhancer-gene pair.

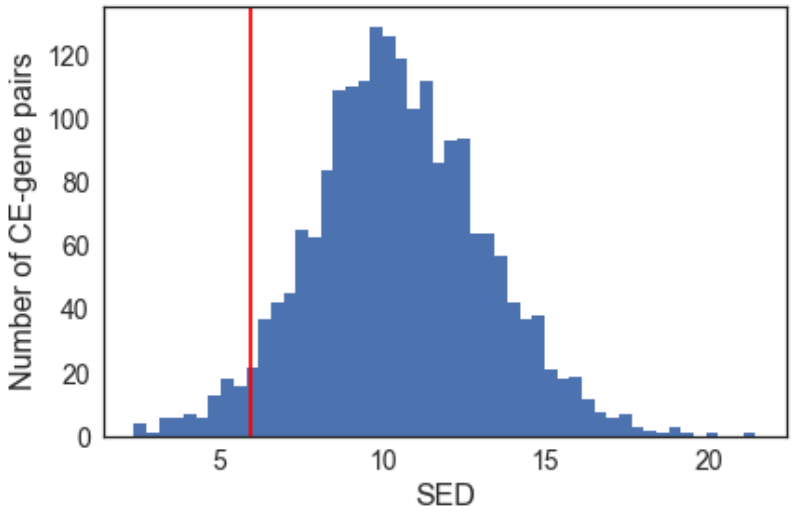


Figure 48 – Histogram of SEDs for Foxh1 candidate gene targets.
Highlighted in red is the SED value for the correlation between *mix1* expression and p300 binding dynamics at the gene's promoter and at a nearby p300-Foxh1 candidate enhancer. CE – Candidate enhancer.

Candidate target genes with low SED values (lowest 30% SED values) are 1.5 times more likely to have decreased expression (>50% fold change) in the FoxH1 morpholino knockdown study than genes with high SED values, however, this was not statistically significant (p-value = 0.12, Fisher Exact test).

5.4 Discussion

p300 seems to be involved in the patterning of all germ layers and their development, by interacting with several essential transcription factors. Most of the identified transcription factor families were known to interact with p300, however, to my knowledge, there has not been any reports of p300-Zic interactions. It would be interesting to experimentally validate this and potentially expand the set of known p300 interaction partners. The differential motif binding analysis also provided an approach to predict enhancers involved in the different developmental processes.

I then investigated the prediction of candidate gene targets for transcription factors, based on the p300 association with their motifs at different time points, using Foxh1 as proof of concept.

Selecting p300-Foxh1 candidate enhancer regions yielded a set of sequences which matched extremely well with previously published Foxh1 ChIP-seq data (99.5% of p300-Foxh1 regions were also detected in that study). However, only 3% of the previously published Foxh1 regions were detected using this method. A reason for this low number may be that not all Foxh1 binding in the genome are associated with p300. However, it is likely mainly due to the peak caller parameters, which in this study were set quite stringently, and have an extremely high impact on how many regions are called. In the present study, only 17,414 p300 regions were detected, therefore the number of p300-Foxh regions would always be a small fraction of the total Foxh1 regions detected by Charney and colleagues (40,884).

I then calculated the SED between each of the p300-Foxh1 candidate enhancers and any gene within 1 Mb, in order to predict candidate target genes. The top hit, *cdk9*, was not affected by *foxh1* knockdown in previously published

data (Chiu et al., 2014), as well as seven of the top 10 hits. This study only had one replicate and imposed a fold change threshold of 1.5. Some of these candidate target genes may in fact be Foxh1 targets but having fold changes smaller than the applied threshold. In another ongoing study, Nick Owens reanalysed the Chiu et al., 2014 data and *cdk9* is indeed down regulated, with a sub-threshold fold change of 1.25. A confounding factor is also that some transcription factors are redundant and their function can be in part replaced, therefore knockdown experiments have high percentages of false negatives (Dai et al., 2009, Gitter et al., 2009, Wu and Lai, 2015).

This analysis has allowed the investigation of many possible enhancer-transcription factor-gene trios with a single dataset. Whilst a specific transcription factor ChIP-seq combined with gene knockdown would be more accurate for a given transcription factor, this provides a much broader picture without the need to perform the experiments for every transcription factor of interest. This data can then be used by researchers to focus on a small number of candidate regions that can be validated by transcription factor specific ChIP-qPCR.

Chapter 6. Discussion

With the advent of next-generation sequencing and with decreasing costs, genome-wide studies are now commonplace to elucidate many cellular processes, including, but not limited to, determining where specific proteins bind, where histone modifications are present, and what is the transcriptional state of cells. Few studies, however, have time-resolved data with more than two or three time points. These can be useful to study relatively static processes, however, as it was shown by Collart et al, 2014 and Owens et al, 2016, transcription in the early embryo is highly dynamic and analysis at sparse time points will not capture a full accurate picture of the process under study.

Given the significant dynamics of transcription in the early embryo, the mechanisms regulating transcription should also display dynamic behaviours. p300 is widely used to predict active enhancer (Wang et al., 2005, Heintzman et al., 2007, Heintzman et al., 2009, Visel et al., 2009a, Wang et al., 2009, Blow et al., 2010, May et al., 2011, Rada-Iglesias et al., 2011, Attanasio et al., 2013) and Hontelez et al., 2015 indicated that its binding changes during early *X. tropicalis* development, however that study only determined p300 binding at five time points over a period of about 24 hours. If p300 binding is as dynamic as transcription, those time intervals would be insufficient to correctly calculate enhancer usage dynamics, as determined by p300 binding. Therefore, in this project I set out to determine the dynamics of enhancer usage and if/how they correlate with gene transcription.

6.1 Candidate enhancers and p300 dynamics

In the work presented in the third chapter of this thesis I aimed to create a dataset of candidate enhancers and determine how p300 binding dynamics change in these regions. This was achieved by generating and analysing two high resolution p300 ChIP-seq time series.

6.1.1 p300 ChIP-seq analysis

I started by optimising the p300 ChIP-seq protocol, particularly focusing on the details which can impact time series studies. This was discussed in Chapter 3, therefore I will only focus on the most important step in this optimisation process: The normalisation method. This was essential for a quantitative analysis of ChIP-seq time series data, not dependent on varying efficiencies in immunoprecipitation resulting in different levels of background signal between different samples. This normalisation implicitly assumes that total p300 binding remains constant in the embryo during the times sampled. Whilst this assumption is not ideal, it is the assumption used to normalise expression in most RNA-seq studies. Moreover, it should not affect the temporally local dynamics at a given enhancer, if the total level of p300 binding varies more slowly than that at a given enhancer. Given that all time points assayed are post-MBT, the relative stability of p300 mRNA (Figure 5), and that attention has been restricted to high SNR, variable (> 2-fold change in occupancy) p300 binding, this is not an unreasonable assumption and normalisation.

6.1.2 p300 binding dynamics

A dataset of 12,572 p300 regions was created, with varied binding dynamics between 5 and 17.5 hpf, an interval which encompasses the end of the blastula stages, gastrulation and most of neurulation, some of the most important stages of development. From these, 9,807 are intergenic or intronic, being candidate enhancers. This is a useful tool for *Xenopus* researchers, allowing the fine-tuning of gene expression by disrupting enhancer regions and also to potentially study how mutations in non-coding regulatory regions can lead to diseases and altered phenotypes.

I showed that p300 binding is extremely dynamic, which was an important finding, as the overall hypothesis of this project was that, given that transcription in the early embryo if highly dynamic, the underlying regulatory mechanisms should also present dynamic behaviours. With such a highly dynamic binding, this data also showed the importance of performing experiments at a high temporal resolution. By assaying development at lower resolution these fast dynamics would be lost.

p300 binding at candidate enhancers was shown to be more dynamic than in promoters (discussed in 3.3.2 – Long time series). This finding reinforced the model of several enhancers regulating the same promoter, at different time points and different cell types.

6.2 p300 and transcription

In the work presented in the fourth chapter of this thesis I aimed to understand how p300 binding and transcription dynamics correlate, by making use of the p300 data generated and published RNA-seq data. I specifically set out to determine if active genes are more likely to have p300 binding nearby; if the p300 binding at promoters and nearby enhancers had similar dynamics; if p300 binding and gene transcription had similar dynamics; and finally, if an enhancer-gene pairing method could be developed.

6.2.1 p300 and active genes

I showed that active genes are more likely to have p300 binding in their promoter and nearby regions than inactive genes, and that p300 occupancy was higher in promoters of active than inactive genes, which is coherent with the model of p300 being involved in transcription activation. Furthermore, the more highly activated a gene is, the more likely it is to have p300 binding and the p300 occupancy at its promoter is also higher, compared to less activated genes.

6.2.2 p300 binding at promoters and candidate enhancers

If p300 is indeed involved in enhancer-promoter looping, it would be expected that p300 binding dynamics in both regions involved in the same loop would be similar, given that the same molecule is in close proximity to both regions. As previously mentioned, it is challenging to predict which enhancer regulates which genes, therefore I initially restricted this analysis to small genomic distances (20 kb) and determined that p300 binding at candidate enhancers and promoters

was indeed more correlated than randomly generated pairs. This reinforces the model of p300 being involved in enhancer-promoter looping.

6.2.3 p300 binding and gene transcription

I also showed that p300 binding at proximal promoters correlates with gene transcription levels. This reinforces the model that enhancer-promoter looping is involved in transcriptional regulation and suggests that the loop is maintained for the duration of transcription, otherwise, p300 dynamics would only correlate with transcription activation, and not with the entire gene expression profile.

6.2.4 Candidate enhancer-gene pairing

Having shown that p300 binding dynamics at promoters correlates with both p300 dynamics at candidate enhancers and with gene expression, I set out to develop an enhancer-gene pairing method.

Currently, there is no reliable way to determine which enhancer regulates which gene in a genome-wide manner. Chromatin conformation capture techniques can be useful to determine these enhancer-gene pairs for individual cases, but applications of these approaches on a genome-wide scale lack resolution (reviewed by de Wit and de Laat, 2012, Denker and de Laat, 2016). Several studies (for example, Ernst et al., 2011, Thurman et al., 2012, Corradin et al., 2014, Cao et al., 2017) have attempted to pair genes and their corresponding enhancers, mainly in human cells, assisted by the vast amounts of data generated for these cells. For example, Ernst and colleagues generated epigenome maps for nine histone marks in nine cell types (Ernst et al., 2011) and Cao and colleagues made use of published data on seven different histone modifications, DNAse I, DNA

methylation, eRNAs, ChIA-PET and Hi-C for 935 samples (Cao et al., 2017) to predict enhancer-gene pairs in human cells.

Using such rich datasets are the ideal way to computationally predict enhancer-gene pairs, however, with the associated costs, these approaches are limited to genomes such as the human or mouse. Therefore, I developed a method – Sum of Euclidean Distances (SED) – to predict candidate enhancer-gene pairs based on p300 ChIP-seq and RNA-seq time series data. The advantages, limitations and possible solutions were discussed in 4.8 – Discussion.

This method led to the prediction that only 25% of genes are regulated by the closest enhancer. As discussed previously, this is an important finding, given that most ChIP-seq regions annotation algorithms uses distance as their main or sole parameter. If this SED method is validated, it will in the least define error boundaries for the closest enhancer heuristic and it will also suggest ways in which these conventional approaches may be improved.

6.2.5 Enhancer RNAs

Taking advantage of the set of candidate enhancers generated in this study and the available RNA-seq data from Owens et al, 2016, I showed that potential eRNAs correlate well with p300 binding in the same region and with nearby gene expression. This was the first time potential eRNAs were reported in *Xenopus* and almost 10% of them may be maternally deposited. There are several confounding issues associated with eRNA analysis, discussed in Chapter 4, however it would be interesting to pursue further studies on this topic, particularly as it may provide an extra parameter to predict enhancer-gene pairs.

6.2.6 Genome annotation

When filtering the p300 regions to analyse potential eRNAs, I realised that 36% of regions annotated at intergenic had H3K4me3 peaks (as detected in Hontelez et al., 2015). This histone modification is extensively used to predict promoter regions (Encode et al., 2007, Heintzman et al., 2007). This high percentage of H3K4me3 in regions annotated as being intergenic may be due to either incorrect gene models/unannotated genes, or due to H3K4me3 not being exclusively present in promoter regions. The actual cause is likely a combination of the two, with some H3K4me3 representing alternative promoters or promoters of unannotated genes, and some representing non-promoter H3K4me3. H3K4me3 ChIP-seq data should be combined with RNA-seq data to attempt to distinguish between the two, leading to a more accurate list of genes and their alternative promoters, which is an essential tool for *Xenopus* researchers in the genomic era.

6.3 p300 differential motif analysis

In the work presented in Chapter 5 I aimed to investigate p300 binding in *X. tropicalis* development and to determine if the data generated in this project could be used to predict candidate target genes for a diverse set of transcription factors.

I identified which DNA binding motifs were associated with p300 occupancy at different developmental stages and, for all motifs, at least some members of each transcription factor family had an expression profile highly concordant with the motif association with p300. These motifs are bound by transcription factors involved in very diverse developmental processes, indicating that p300 may also be involved in regulating those processes, however, even more important is the indication that p300 activity and motif analysis can be coupled to predict transcription factor binding and potential target genes. Therefore, I attempted to predict the target genes for a given transcription factor, based on the p300 binding association with its motif and the enhancer-gene pairing method.

The method described on Chapter 5 allows the prediction of candidate target genes for multiple transcription factors with a single dataset. This allows researchers to then validate a small number of candidate enhancers by ChIP-qPCR for the transcription factor of interest, followed by reporter assays (e.g Tol2-mediated transgenesis) or enhancer mutation/deletion (e.g CRISPR-Cas9) to validate enhancer function and target gene.

6.4 Future work

The following are the most interesting paths I believe should be followed as a continuation of the work described in this thesis:

1. In this project I analysed p300 binding in whole embryos, thus, the obtained results represent averages of different processes and gene expression programs occurring in different tissues. It would be interesting to perform p300 ChIP-seq time series in specific tissues. This could be achieved, for example, by adapting BiTS-ChIP to *Xenopus*. Batch isolation of tissue-specific chromatin for Immunoprecipitation (BiTS-ChIP) was developed by Bonn and colleagues in *D. melanogaster*. In this technique embryos need to express a cell type specific marker to allow FACS sorting of the fixed nuclei, which are then processed with the standard ChIP-seq protocol (Bonn et al., 2012).
2. Determine if there is a time delay between p300 binding and gene activation. This could be achieved by performing the ED analysis with the RNA-seq data delayed by X number of hours and determine if a given time delay led to a higher correlation.
3. 3C or 4C to validate some of the predicted enhancer-promoter interactions. Low SED enhancer-promoter-gene would be the most interesting trios to test initially.
4. CRISPR-mediated knockout of candidate enhancer regions in predicted enhancer-promoter pairs to determine if gene expression is reduced.

5. I showed that p300 binding at candidate enhancers is more dynamic than at promoters, likely due to genes being regulated by multiple enhancers. Therefore, it would be interesting to assess if it is possible to determine pairs of enhancers which together explain the dynamics of p300 binding at a given promoter.
6. Further work on eRNAs, attempting to solve the confounding issues and potentially use eRNA data to improve the enhancer-gene predictions.
7. An interesting extension to the analysis described in Chapter 5 would be to investigate the co-occurrence of motifs within a p300 region, to potentially uncover co-factors or synergistic partners.

6.5 Conclusion

In this thesis, I described the generation of a candidate enhancer dataset and their usage in early *X. tropicalis* development, using p300 as a proxy. I then showed that p300 binding at candidate enhancers and promoters correlate, as well as p300 binding in promoters and gene expression. This reinforced the model of p300-mediated looping being involved in gene transcription and indicated that the loop structure is likely maintained during transcription. Finally, I developed a method to predict enhancer-gene pairs and another to predict candidate target genes for a given transcription factor.

These results and tools can be useful for researchers to further understand transcriptional regulation and regulatory networks.

Chapter 7. Appendix

7.1 Appendix 1 – Predicted enhancer-gene pairs

Gene	Chromosome	Start	End	Distance	SED
Xetro.A02181|mknk2	scaffold_1	123650317	123650643	-11266	2.24
Xetro.J00801|gs17	scaffold_10	30778205	30778746	-2468	2.38
Xetro.B02193|fzd4	scaffold_2	133845672	133845971	-19888	2.44
Xetro.G00372|	scaffold_7	16066598	16066914	1398	2.54
Xetro.F00048|eif2c2	scaffold_6	1719646	1719990	70922	2.60
Xetro.E00206|sox2	scaffold_5	12415547	12415804	-57322	2.67
Xetro.H01485|pkdcc.1	scaffold_8	74627972	74628265	76492	2.69
Xetro.H01348|meis3	scaffold_8	65765102	65765419	66936	2.76
Xetro.I00911|	scaffold_9	49357636	49358088	-8686	2.89
Xetro.A01003|sds	scaffold_1	61803635	61803947	-92287	2.89
Xetro.G01863|	scaffold_7	104717612	104717882	-70436	2.90
Xetro.D02282|pim1	scaffold_4	121583870	121584267	67048	2.97
Xetro.D00482|sall1	scaffold_4	22515068	22515221	-7842	3.14
Xetro.I00092|mcm6.2	scaffold_9	3637092	3637359	-58685	3.15
Xetro.J00216|arhgap39	scaffold_10	6161572	6161769	-4795	3.24
Xetro.G01450|hes4	scaffold_7	82683994	82684282	15368	3.27
Xetro.F01821|c3orf54	scaffold_6	137141745	137142166	35534	3.35
Xetro.A01107|lpcat4	scaffold_1	64930391	64930730	-3302	3.41
Xetro.E00971|dll1	scaffold_5	73327772	73328282	3950	3.45
Xetro.B01595|unnamed	scaffold_2	93984726	93985145	-59327	3.46
Xetro.G00640|slc16a12	scaffold_7	32463266	32463558	1287	3.49
Xetro.I02074|msgn1	scaffold_9	96845192	96845681	-10904	3.50
Xetro.H01545|znf238.2	scaffold_8	78174109	78174427	-1680	3.52
Xetro.K04987|vsig8	scaffold_8c	1233561	1234784	-14639	3.56
Xetro.G01634|hes3.3	scaffold_7	94447157	94448064	-5844	3.62
Xetro.H02244|plekho1	scaffold_8	105430922	105431167	75856	3.67
Xetro.A02912|	scaffold_1	173063869	173064439	12522	3.76
Xetro.I00336|pdk1	scaffold_9	18587421	18587778	76966	3.83
Xetro.F00500|cebpd	scaffold_6	36715084	36715261	-65587	3.87
Xetro.H00596|cdx4	scaffold_8	26394038	26394973	-96248	3.89
Xetro.H01053|pbx2	scaffold_8	52222789	52222999	41968	3.91
Xetro.D01106|foxa4	scaffold_4	61589850	61590377	11236	3.96
Xetro.G01635|hes8	scaffold_7	94441545	94441986	10702	3.98
Xetro.K00363|mef2d	scaffold_27	100503	100684	32486	4.01
Xetro.E01636|mixer	scaffold_5	118173284	118173610	13120	4.02
Xetro.B01852|tnfrsf19	scaffold_2	108191436	108191641	24984	4.03
Xetro.A01563|	scaffold_1	86445086	86445275	-43248	4.05
Xetro.F00583|gata6	scaffold_6	41132586	41132959	-2993	4.05

Xetro.D01928|hpdl	scaffold_4	102778658	102778965	70271	4.07
Xetro.B01645|efnb2	scaffold_2	96714948	96715267	40670	4.08
Xetro.G02292|mrto4	scaffold_7	123731987	123732226	-85314	4.20
Xetro.H02253|nr2f5	scaffold_8	105875536	105875894	17801	4.28
Xetro.A01542|efs	scaffold_1	85947340	85947646	16445	4.32
Xetro.A00551|znf608	scaffold_1	38115300	38115793	1688	4.33
Xetro.C00582|myef2	scaffold_3	19938489	19938791	-8420	4.33
Xetro.D01817|	scaffold_4	94644338	94644672	9953	4.34
Xetro.G01845|tead4	scaffold_7	104002689	104003619	36260	4.36
Xetro.J00553|dlx3	scaffold_10	21491713	21491894	22952	4.37
Xetro.A01061|aplnr	scaffold_1	64036047	64036477	3885	4.38
Xetro.I01918|	scaffold_9	89274927	89275203	5838	4.39
Xetro.F01630|fzd8	scaffold_6	124211435	124212040	2554	4.39
Xetro.A01622|gas1	scaffold_1	90565037	90565940	-5466	4.42
Xetro.G00717|admp2	scaffold_7	36586207	36586723	7626	4.43
Xetro.B01266|rnd1	scaffold_2	77680506	77680725	-34916	4.44
Xetro.C01640|btg1	scaffold_3	84535572	84535800	4506	4.45
Xetro.C01899|	scaffold_3	102962854	102963084	54094	4.47
Xetro.I01389|unnamed	scaffold_9	77936317	77936543	47902	4.48
Xetro.K01603|	scaffold_129	196861	197219	-14664	4.48
Xetro.C00433|zmiz2	scaffold_3	11318934	11319570	30336	4.51
Xetro.D01357|ptgs2	scaffold_4	75023843	75024324	-5130	4.52
Xetro.A00935|mn1	scaffold_1	57936682	57937079	6947	4.53
Xetro.E01089|hnrnpu	scaffold_5	79663277	79663600	-1048	4.53
Xetro.I01107|emp2	scaffold_9	62689798	62690207	-17126	4.55
Xetro.B01215|	scaffold_2	75288602	75289054	-44832	4.56
Xetro.H01434|bmp4	scaffold_8	72120991	72121226	-1412	4.58
Xetro.H00338|nr6a1	scaffold_8	15960873	15961329	-5053	4.59
Xetro.G00035|rbm20	scaffold_7	1850461	1850948	54879	4.67
Xetro.F01742|ngfr	scaffold_6	133748766	133749077	6673	4.72
Xetro.I02095|prss29	scaffold_9	97565341	97565658	-32338	4.73
Xetro.F00346|unnamed	scaffold_6	25828162	25828377	-94588	4.76
Xetro.I00321|sp5	scaffold_9	17802012	17802348	6004	4.77
Xetro.E01369|flrt3	scaffold_5	98375616	98376185	70605	4.77
Xetro.K05077|anxa9	scaffold_8c	4142379	4142676	-8845	4.79
Xetro.C01387|msx2	scaffold_3	70954158	70954516	71620	4.80
Xetro.A00953|unnamed	scaffold_1	59228944	59229330	20541	4.81
Xetro.I02106|prss27	scaffold_9	97705804	97706130	33321	4.81
Xetro.G01448|plekhn1	scaffold_7	82475104	82475331	-8270	4.85
Xetro.A01113|	scaffold_1	64930391	64930730	55042	4.87
Xetro.D00417|	scaffold_4	20092946	20093488	2317	4.88
Xetro.J00830|rara	scaffold_10	31996315	31996521	71264	4.89
Xetro.K05138|gata4	scaffold_5b	1541015	1541460	-7060	4.91
Xetro.D00195|	scaffold_4	9574649	9575070	-2331	4.93

Xetro.E01541|atf3	scaffold_5	112283306	112283625	-99947	4.94
Xetro.A02050|klb	scaffold_1	119305852	119306066	98234	4.94
Xetro.G00875|	scaffold_7	46702957	46703221	11567	4.97
Xetro.A00363|isl1	scaffold_1	22750517	22750847	-1797	4.98
Xetro.C00277|	scaffold_3	7849224	7849516	-17540	4.98
Xetro.G01800|slc6a16	scaffold_7	102348618	102348950	83486	5.01
Xetro.H00049|	scaffold_8	2090074	2090296	-6681	5.03
Xetro.A00602|xbp1	scaffold_1	42689082	42689305	86038	5.06
Xetro.E00316|slc33a1	scaffold_5	20923348	20923573	-45134	5.10
Xetro.H00248|znf750	scaffold_8	9906228	9906402	14815	5.12
Xetro.K00691|	scaffold_40	428625	428813	-4609	5.14
Xetro.F00226|fzd6	scaffold_6	18333997	18334273	2108	5.15
Xetro.A03177|msx1	scaffold_1	193646007	193646727	-5191	5.18
Xetro.D01662|	scaffold_4	85654012	85654311	23476	5.21
Xetro.F00234|grhl2	scaffold_6	18937961	18938192	1904	5.23
Xetro.D00899|cdh1	scaffold_4	52535949	52536239	13043	5.23
Xetro.G00599|pcdh8l	scaffold_7	30280138	30280795	-13206	5.25
Xetro.D00013|	scaffold_4	637801	638158	47903	5.26
Xetro.A02603|tet2	scaffold_1	148894084	148894252	20634	5.26
Xetro.J00920|cass4	scaffold_10	35377485	35377687	94683	5.27
Xetro.F00849|foxc1	scaffold_6	63060924	63061147	-3010	5.33
Xetro.C00618|gatm	scaffold_3	21308488	21308700	-3152	5.39
Xetro.A02254|cnn1	scaffold_1	126491892	126492134	-92205	5.39
Xetro.K01222|xnf7	scaffold_83	382217	382460	-13346	5.39
Xetro.C01366|afap1l1	scaffold_3	69115012	69115231	-75259	5.40
Xetro.A01974|rod1	scaffold_1	116646417	116646710	41766	5.41
Xetro.A01009|tbx3	scaffold_1	62144303	62144581	-1270	5.45
Xetro.F01133|bmper	scaffold_6	87720271	87720744	2470	5.45
Xetro.D00701|kctd15	scaffold_4	38245268	38245657	36140	5.47
Xetro.B01533|rarg	scaffold_2	91515818	91516206	-5506	5.47
Xetro.B00282|tceb3	scaffold_2	14054792	14055389	-31128	5.48
Xetro.B00508|asb9	scaffold_2	26346879	26347298	-3929	5.50
Xetro.B02279|	scaffold_2	139170699	139170887	-38343	5.53
Xetro.A01550|slc7a8	scaffold_1	86138288	86138554	-17177	5.53
Xetro.D00441|	scaffold_4	20886930	20887459	3206	5.55
Xetro.C00514|tet3	scaffold_3	14213073	14213542	-8561	5.56
Xetro.E00193|hes1	scaffold_5	11302709	11302907	1431	5.56
Xetro.D01445|rbfox2	scaffold_4	79974701	79975104	8032	5.58
Xetro.G01018|anxa2	scaffold_7	53249145	53249319	80976	5.63
Xetro.I01474|	scaffold_9	80330421	80330704	-2667	5.66
Xetro.A03291|foxi4.2	scaffold_1	201659904	201660438	-10866	5.66
Xetro.K01424|lpar2	scaffold_109	44823	45127	-31315	5.69
Xetro.B00724|srsf1	scaffold_2	40724349	40724501	77330	5.69
Xetro.D02243|gata2	scaffold_4	120113071	120113225	42082	5.70

Xetro.H01785|meis2	scaffold_8	90219071	90219299	-98555	5.70
Xetro.D01086|tmed10	scaffold_4	60985721	60986011	-79031	5.70
Xetro.J00091|srsf2	scaffold_10	3664309	3664592	-94667	5.72
Xetro.E00553|ripply2.1	scaffold_5	40771280	40771569	-20420	5.75
Xetro.J00381|snai1	scaffold_10	12981662	12981883	-1065	5.76
Xetro.A00618|	scaffold_1	43088040	43088184	15158	5.77
Xetro.G01539|	scaffold_7	88943004	88943216	-82740	5.80
Xetro.I00468|fzd7	scaffold_9	27336503	27336905	-3082	5.82
Xetro.B00654|spns2	scaffold_2	36468647	36469013	-4125	5.84
Xetro.E00181|chrd	scaffold_5	10692051	10692208	2218	5.85
Xetro.C00636|mex3b	scaffold_3	22486553	22486912	25166	5.88
Xetro.A02274|loc388630	scaffold_1	127237573	127237925	15765	5.91
Xetro.H00544|p2ry4	scaffold_8	24340264	24340496	-10470	5.93
Xetro.C01869|szl	scaffold_3	100190388	100190772	-5664	5.94
Xetro.B02316|rps3	scaffold_2	140829491	140829751	19273	5.95
Xetro.B01113|a2ld1	scaffold_2	64808938	64809143	76392	5.98
Xetro.B01496|hnrnpa1	scaffold_2	89788156	89788319	-3724	5.98
Xetro.E01129|akap12	scaffold_5	82206815	82206995	5649	6.01
Xetro.F01555|spag6	scaffold_6	118702130	118702325	-35846	6.01
Xetro.H01766|srsf5	scaffold_8	89005510	89005751	-2546	6.03
Xetro.A02777|pcdh18	scaffold_1	162559481	162559769	6277	6.03
Xetro.D00077|scube2	scaffold_4	3349289	3349814	-10469	6.07
Xetro.F00764|rbm24	scaffold_6	57096034	57096331	61008	6.13
Xetro.A01783|utp15	scaffold_1	101933202	101933429	49924	6.16
Xetro.F01813|sema3f	scaffold_6	136745058	136745490	2790	6.17
Xetro.J00103|cygb	scaffold_10	3766521	3766867	-8642	6.20
Xetro.G00462|ventx3.2	scaffold_7	21504204	21504370	-54461	6.24
Xetro.B00695|ksr1	scaffold_2	38542330	38542479	-70638	6.25
Xetro.C00179|tacc1	scaffold_3	5761661	5761998	-6932	6.26
Xetro.A00753|fzd10	scaffold_1	49748767	49748922	18821	6.27
Xetro.H00800|	scaffold_8	43654950	43655148	15035	6.29
Xetro.K00983|ahnak	scaffold_65	338506	338799	-49347	6.29
Xetro.D01314|atf4	scaffold_4	73577515	73577750	-1072	6.29
Xetro.K05002|tagln2	scaffold_8c	1464034	1464216	39825	6.29
Xetro.D01032|prpf39.2	scaffold_4	58857552	58857832	81153	6.31
Xetro.K00373|fkbp9	scaffold_27	340557	340961	18827	6.34
Xetro.I00651|hes6.1	scaffold_9	35394488	35394987	6850	6.37
Xetro.A02550|rasgef1b	scaffold_1	144836022	144836203	-61922	6.38
Xetro.H01321|axl	scaffold_8	64902347	64902629	-29799	6.43
Xetro.F00287|	scaffold_6	21901715	21902064	63888	6.43
Xetro.A03043|kit	scaffold_1	183676655	183677109	-2465	6.43
Xetro.J00905|bmp7.1	scaffold_10	34842324	34842935	-70130	6.46
Xetro.C00970|cdx1	scaffold_3	44797719	44798183	-9311	6.46
Xetro.I00194|zeb2	scaffold_9	7328087	7328488	67048	6.47

Xetro.K00372|rab25	scaffold_27	332361	332701	8228	6.47
Xetro.G01640|hes3.1	scaffold_7	94488011	94488497	29486	6.48
Xetro.H00955|degs3	scaffold_8	50045360	50045588	-84024	6.48
Xetro.J00368|eya2	scaffold_10	11858258	11858640	-42497	6.54
Xetro.I00863|arl4c	scaffold_9	46888958	46889320	-9895	6.57
Xetro.G00498|nt5c2	scaffold_7	23389749	23389913	-74311	6.58
Xetro.A02917|	scaffold_1	173393370	173393721	1920	6.59
Xetro.E00432|fam83b	scaffold_5	32172480	32172888	-4968	6.64
Xetro.B01377|pcdh8.2	scaffold_2	83178240	83178723	-33523	6.65
Xetro.K02114|cdh26	scaffold_217	44405	44676	-1522	6.66
Xetro.H01276|serpina1	scaffold_8	62420822	62421193	-92840	6.70
Xetro.F01674|ptprn2	scaffold_6	128512550	128512814	26022	6.70
Xetro.C02032|gata3	scaffold_3	111971032	111971243	90060	6.74
Xetro.I02069|nt5c1a	scaffold_9	96604687	96604939	-21487	6.74
Xetro.A01750|ccdc125	scaffold_1	100912509	100912816	71702	6.75
Xetro.A01688|foxd4l1.1	scaffold_1	95945615	95945810	20956	6.76
Xetro.A01618|fbp1	scaffold_1	90318546	90318896	-58879	6.77
Xetro.F00798|tfap2a	scaffold_6	59748726	59748889	16662	6.79
Xetro.E00631|prdm1	scaffold_5	48169600	48170036	-1858	6.79
Xetro.H00776|pou4f1.2	scaffold_8	42275685	42276076	40234	6.81
Xetro.I00090|cxcr4	scaffold_9	3506700	3506936	-38208	6.87
Xetro.A02297|slc25a42	scaffold_1	128242860	128243146	69043	6.87
Xetro.C00695|mespb	scaffold_3	25211058	25211457	-26774	6.90
Xetro.D00916|nudt22	scaffold_4	53508614	53509057	-3316	6.92
Xetro.K02331|ddr2	scaffold_266	79908	80646	-1873	6.92
Xetro.E01382|sptlc3	scaffold_5	99150963	99151266	83238	6.95
Xetro.K00375|ca14	scaffold_27	439106	439612	-8842	7.00
Xetro.I00424|idh1	scaffold_9	24441640	24441906	73695	7.01
Xetro.A02690|lef1	scaffold_1	155883080	155883362	-13429	7.04
Xetro.B00284|eef1a1o	scaffold_2	13978101	13978311	76884	7.08
Xetro.E01547|vash2	scaffold_5	112283306	112283625	60021	7.10
Xetro.C01646|angpt4	scaffold_3	84955772	84955929	13461	7.12
Xetro.K00255|s1pr5	scaffold_24	267889	268320	-89132	7.14
Xetro.B00797|gpr143	scaffold_2	45166838	45167226	-12160	7.17
Xetro.B01816|cdx2	scaffold_2	106282766	106283241	9286	7.17
Xetro.A00284|ankdd1b	scaffold_1	17472707	17472984	3062	7.19
Xetro.I01555|gng13	scaffold_9	83360861	83361207	36301	7.20
Xetro.B02008|mmp3	scaffold_2	120686664	120686900	-29984	7.22
Xetro.G01589|unnamed	scaffold_7	91197760	91198088	76583	7.22
Xetro.B00091|grhl3	scaffold_2	4077552	4077776	-80840	7.24
Xetro.E00740|sgk1	scaffold_5	55823951	55824302	2368	7.32
Xetro.H01277|gsc	scaffold_8	62328059	62328276	92840	7.34
Xetro.A00398|c9	scaffold_1	25871695	25872228	3069	7.34
Xetro.A01289|unnamed	scaffold_1	73734177	73734614	-44632	7.38

Xetro.B00414|upk3b	scaffold_2	21245416	21245648	4046	7.39
Xetro.H01281|clmn	scaffold_8	62725770	62726111	-69797	7.39
Xetro.D00831|chst6	scaffold_4	47969568	47969728	95652	7.40
Xetro.G00241|ank3	scaffold_7	9449257	9450246	-14476	7.41
Xetro.D01570|	scaffold_4	82495772	82496206	18811	7.42
Xetro.K02417|atp1b2	scaffold_297	2357	2688	28606	7.43
Xetro.G01742|eps8l1	scaffold_7	99632946	99633184	-6793	7.44
Xetro.H00296|unnamed	scaffold_8	13300745	13300903	-81121	7.46
Xetro.E01409|capn8	scaffold_5	101822720	101823049	-24432	7.47
Xetro.D02572|fgd1	scaffold_4	139918482	139918847	84460	7.48
Xetro.I00143|tuba3e	scaffold_9	5275055	5275677	-15690	7.49
Xetro.E00860|ezr	scaffold_5	65035333	65035563	11062	7.50
Xetro.H02267|	scaffold_8	106743108	106743614	-28486	7.50
Xetro.B01525|sp7	scaffold_2	91229553	91229811	5149	7.50
Xetro.J00120|evpl	scaffold_10	4064214	4064649	91942	7.50
Xetro.D02526|hesx1	scaffold_4	137537006	137537356	-12651	7.50
Xetro.C01851|atp6ap1	scaffold_3	98711128	98711441	7882	7.53
Xetro.B01543|krt8	scaffold_2	91915507	91915812	58611	7.64
Xetro.A02998|fat1	scaffold_1	180347698	180348340	6271	7.66
Xetro.E00345|mycn	scaffold_5	23536773	23536954	-54033	7.66
Xetro.H00927|rtn2	scaffold_8	49191902	49192194	25935	7.75
Xetro.F00785|dsp	scaffold_6	58573329	58573515	-7374	7.76
Xetro.A02845|rps3a	scaffold_1	168147782	168148004	8778	7.79
Xetro.D01614|adap1	scaffold_4	84164599	84164877	-1836	7.81
Xetro.A00866|c22orf36	scaffold_1	55403032	55403245	-68890	7.84
Xetro.G02071|trpm5	scaffold_7	115241191	115241517	-1086	7.84
Xetro.D01090|efemp2	scaffold_4	60906709	60906961	79031	7.90
Xetro.I02104|	scaffold_9	97595993	97596291	97470	7.93
Xetro.E01338|meis1	scaffold_5	95545302	95545551	11152	7.98
Xetro.F00600|lpin2	scaffold_6	42936516	42936815	-32661	7.98
Xetro.E00241|slc2a2	scaffold_5	15989746	15990126	-1859	8.03
Xetro.F01109|gli3	scaffold_6	85801901	85802133	50856	8.04
Xetro.B00360|rab34	scaffold_2	17784970	17785321	36154	8.06
Xetro.I01508|	scaffold_9	81281457	81281732	2352	8.07
Xetro.D01275|cfh	scaffold_4	71316373	71316644	98231	8.09
Xetro.E00020|plod2	scaffold_5	1774690	1775021	-5848	8.14
Xetro.K00087|p2ry11	scaffold_18	613045	613441	-3573	8.18
Xetro.A02031|cpamd8	scaffold_1	118984582	118985031	-32491	8.19
Xetro.E00325|p2ry1	scaffold_5	21617442	21617812	82946	8.19
Xetro.E01282|foxa2	scaffold_5	91327181	91327633	4651	8.20
Xetro.B01312|slc39a5	scaffold_2	80811164	80811418	-98307	8.21
Xetro.B01556|krt5.7	scaffold_2	92300516	92300811	1695	8.24
Xetro.F00712|	scaffold_6	52193749	52194072	-7750	8.25
Xetro.A01296|slc25a22	scaffold_1	73734177	73734614	98340	8.29

Xetro.D00746|aprt	scaffold_4	43308227	43308434	43893	8.31
Xetro.H00155|hmcn2	scaffold_8	6510319	6510517	6884	8.36
Xetro.H01609|six1	scaffold_8	81400727	81401135	-5661	8.37
Xetro.G02246|gramd1a	scaffold_7	121659521	121659735	-41198	8.39
Xetro.B00214|khdrbs1	scaffold_2	10623485	10623922	71143	8.39
Xetro.A02002|hopx	scaffold_1	117775731	117776029	78969	8.44
Xetro.I01651|unnamed	scaffold_9	85490140	85490608	18338	8.45
Xetro.G02296|unnamed	scaffold_7	123752790	123753095	-9986	8.50
Xetro.B01904|pros1	scaffold_2	111316791	111317314	20812	8.52
Xetro.B00098|sh3d21	scaffold_2	4299438	4299868	-77691	8.61
Xetro.E00737|myb	scaffold_5	55335772	55336021	-16246	8.66
Xetro.I00787|ikzf2	scaffold_9	42791291	42791593	14709	8.68
Xetro.D02416|unnamed	scaffold_4	128375576	128375739	-80745	8.82
Xetro.H00900|zic3	scaffold_8	48067925	48068342	-11143	8.82
Xetro.A02236|lppr3	scaffold_1	125648871	125649026	40607	8.89
Xetro.H01740|itpka	scaffold_8	87387889	87388123	18082	8.93
Xetro.E00176|ece2	scaffold_5	10497510	10497989	-59592	9.02
Xetro.H00242|ribc1	scaffold_8	9711232	9711776	43188	9.03
Xetro.D01422|pmm2	scaffold_4	78699509	78699787	12892	9.10
Xetro.B01322|rps26	scaffold_2	80947060	80947224	97448	9.11
Xetro.D00430|lsp1	scaffold_4	20383715	20384059	1477	9.17
Xetro.A00337|rai14	scaffold_1	20365624	20365863	74316	9.18
Xetro.E01625|lefty	scaffold_5	117959395	117959550	-26368	9.36
Xetro.E00199|atp13a-like	scaffold_5	11677579	11677862	4814	9.49
Xetro.D01785|rpe65	scaffold_4	93175487	93175649	26078	9.57
Xetro.A01017|rasal1	scaffold_1	62721533	62721722	24256	9.58
Xetro.A02470|cldn23	scaffold_1	137869626	137869940	-37280	9.77
Xetro.C00684|znf710	scaffold_3	24730708	24730967	13620	9.79
Xetro.I01499|unnamed	scaffold_9	80953646	80954051	22726	9.81
Xetro.A00397|dab2	scaffold_1	25874928	25875134	-75483	9.85
Xetro.D01232|unnamed	scaffold_4	69348405	69348717	45247	9.99
Xetro.G01769|loc388564	scaffold_7	101173028	101173329	3343	10.06
Xetro.E01215|srsf7	scaffold_5	87845803	87846222	12071	10.27
Xetro.A00390|c7	scaffold_1	25205653	25205853	-25966	10.35
Xetro.A01497|hnrnpk	scaffold_1	84974256	84974539	-16655	10.46
Xetro.D00122|prr5l	scaffold_4	6802289	6802513	99439	10.56
Xetro.E00310|tiparp	scaffold_5	20651293	20651571	-67214	10.62
Xetro.J00233|rbl1	scaffold_10	6564844	6565099	69922	10.66
Xetro.E00414|cep19	scaffold_5	30241556	30242103	3395	10.77
Xetro.K00483|slc7a7	scaffold_35	122859	123045	48838	10.88
Xetro.D00429|tnni2	scaffold_4	20248373	20248579	85176	10.89
Xetro.D02419|	scaffold_4	128375576	128375739	20654	10.94
Xetro.D01308|slc25a24	scaffold_4	73108497	73108976	-17232	11.05

Xetro.C01685|wif1	scaffold_3	86350301	86350534	12904	11.15
Xetro.A01936|	scaffold_1	114711081	114711433	8815	11.17
Xetro.D02390|	scaffold_4	126954594	126954924	5854	11.34
Xetro.H00096|unnamed	scaffold_8	4199863	4200170	-42346	11.41
Xetro.G00564|marveld1	scaffold_7	27671556	27671770	8772	11.89
Xetro.H01725|foxa1	scaffold_8	86405259	86405422	-46550	11.95
Xetro.B00348|acy3	scaffold_2	17131362	17131682	56262	13.16
Xetro.H00632|rab39b	scaffold_8	28132496	28132795	55602	13.61
Xetro.G00383|dpysl4	scaffold_7	17115717	17116258	-1350	14.37
Xetro.A03175|d4s234e	scaffold_1	193228303	193228704	87740	19.49

Reference List

AKIMARU, H., HOU, D. X. & ISHII, S. 1997. Drosophila CBP is required for dorsal-dependent twist gene expression. Nat Genet, 17, 211-4.

ALBERT, I., MAVRICH, T. N., TOMSHO, L. P., QI, J., ZANTON, S. J., SCHUSTER, S. C. & PUGH, B. F. 2007. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature, 446, 572-6.

ALLEN, B. L. & TAATJES, D. J. 2015. The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol, 16, 155-66.

ALLENDE, M. L., MANZANARES, M., TENA, J. J., FEIJOO, C. G. & GOMEZ-SKARMETA, J. L. 2006. Cracking the genome's second code: enhancer detection by combined phylogenetic footprinting and transgenic fish and frog embryos. Methods, 39, 212-9.

ANDERSSON, R., GEBHARD, C., MIGUEL-ESCALADA, I., HOOF, I., BORNHOLDT, J., BOYD, M., CHEN, Y., ZHAO, X., SCHMIDL, C., SUZUKI, T., NTINI, E., ARNER, E., VALEN, E., LI, K., SCHWARZFISCHER, L., GLATZ, D., RAITHEL, J., LILJE, B., RAPIN, N., BAGGER, F. O., JORGENSEN, M., ANDERSEN, P. R., BERTIN, N., RACKHAM, O., BURROUGHS, A. M., BAILLIE, J. K., ISHIZU, Y., SHIMIZU, Y., FURUHATA, E., MAEDA, S., NEGISHI, Y., MUNGALL, C. J., MEEHAN, T. F., LASSMANN, T., ITOH, M., KAWAJI, H., KONDO, N., KAWAI, J., LENNARTSSON, A., DAUB, C. O., HEUTINK, P., HUME, D. A., JENSEN, T. H., SUZUKI, H., HAYASHIZAKI, Y., MULLER, F., FORREST, A. R. R., CARNINCI, P., REHLI, M. & SANDELIN, A. 2014. An atlas of active enhancers across human cell types and tissues. Nature, 507, 455-461.

ARANY, Z., NEWSOME, D., OLDREAD, E., LIVINGSTON, D. M. & ECKNER, R. 1995. A family of transcriptional adaptor proteins targeted by the E1A oncoprotein. Nature, 374, 81-4.

ARCECI, R. J., KING, A. A., SIMON, M. C., ORKIN, S. H. & WILSON, D. B. 1993. Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. Mol Cell Biol, 13, 2235-46.

ARNOLD, C. D., GERLACH, D., STELZER, C., BORYN, L. M., RATH, M. & STARK, A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science, 339, 1074-7.

ATTANASIO, C., NORD, A. S., ZHU, Y., BLOW, M. J., LI, Z., LIBERTON, D. K., MORRISON, H., PLAJZER-FRICK, I., HOLT, A., HOSSEINI, R., PHOUANENAVONG, S., AKIYAMA, J. A., SHOUKRY, M., AFZAL, V., RUBIN, E. M., FITZPATRICK, D. R., REN, B., HALLGRIMSSON, B., PENNACCHIO, L. A. & VISEL, A. 2013. Fine tuning of craniofacial morphology by distant-acting enhancers. Science, 342, 1241006.

AYYANATHAN, K., LECHNER, M. S., BELL, P., MAUL, G. G., SCHULTZ, D. C., YAMADA, Y., TANAKA, K., TORIGOE, K. & RAUSCHER, F. J., 3RD 2003. Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev, 17, 1855-69.

BALDWIN, K. M., HADDAD, F., PANDORF, C. E., ROY, R. R. & EDGERTON, V. R. 2013. Alterations in muscle mass and contractile phenotype in response to unloading models: role of transcriptional/pretranslational mechanisms. Front Physiol, 4, 284.

BANERJI, J., OLSON, L. & SCHAFFNER, W. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell, 33, 729-40.

BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell, 27, 299-308.

BANNISTER, A. J., ZEGERMAN, P., PARTRIDGE, J. F., MISKA, E. A., THOMAS, J. O., ALLSHIRE, R. C. & KOUZARIDES, T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature, 410, 120-4.

BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. 2007. High-resolution profiling of histone methylations in the human genome. Cell, 129, 823-37.

BELL, A. C., WEST, A. G. & FELSENFELD, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell, 98, 387-96.

BESSA, J., TENA, J. J., DE LA CALLE-MUSTIENES, E., FERNANDEZ-MINAN, A., NARANJO, S., FERNANDEZ, A., MONTOLIU, L., AKALIN, A., LENHARD, B., CASARES, F. & GOMEZ-SKARMETA, J. L. 2009. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. Dev Dyn, 238, 2409-17.

BLAT, Y. & KLECKNER, N. 1999. Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. Cell, 98, 249-59.

BLOW, M. J., MCCULLEY, D. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., BRISTOW, J., REN, B., BLACK, B. L., RUBIN, E. M., VISEL, A. & PENNACCHIO, L. A. 2010. ChIP-Seq identification of weakly conserved heart enhancers. Nat Genet, 42, 806-10.

BONN, S., ZINZEN, R. P., PEREZ-GONZALEZ, A., RIDDELL, A., GAVIN, A. C. & FURLONG, E. E. 2012. Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. Nat Protoc, 7, 978-94.

BOSE, D. A., DONAHUE, G., REINBERG, D., SHIEKHATTAR, R., BONASIO, R. & BERGER, S. L. 2017. RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell, 168, 135-149 e22.

BOYER, L. A., PLATH, K., ZEITLINGER, J., BRAMBRINK, T., MEDEIROS, L. A., LEE, T. I., LEVINE, S. S., WERNIG, M., TAJONAR, A., RAY, M. K., BELL, G. W., OTTE, A. P., VIDAL, M., GIFFORD, D. K., YOUNG, R. A. & JAENISCH, R. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature, 441, 349-53.

BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008. High-resolution mapping and characterization of open chromatin across the genome. Cell, 132, 311-22.

BRACKEN, A. P., DIETRICH, N., PASINI, D., HANSEN, K. H. & HELIN, K. 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev, 20, 1123-36.

BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods, 10, 1213-8.

BUENROSTRO, J. D., WU, B., CHANG, H. Y. & GREENLEAF, W. J. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol, 109, 21 29 1-9.

BULGER, M. & GROUDINE, M. 2011. Functional and mechanistic diversity of distal transcription enhancers. Cell, 144, 327-39.

CAO, Q., ANYANSI, C., HU, X., XU, L., XIONG, L., TANG, W., MOK, M. T. S., CHENG, C., FAN, X., GERSTEIN, M., CHENG, A. S. L. & YIP, K. Y. 2017. Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. Nat Genet, 49, 1428-1436.

CAO, Y., KNOCHEL, S., DONOW, C., MIETHE, J., KAUFMANN, E. & KNOCHEL, W. 2004. The POU factor Oct-25 regulates the Xvent-2B gene and counteracts terminal differentiation in Xenopus embryos. J Biol Chem, 279, 43735-43.

CAO, Y., SIEGEL, D. & KNOCHEL, W. 2006. Xenopus POU factors of subclass V inhibit activin/nodal signaling during gastrulation. Mech Dev, 123, 614-25.

CARROZZA, M. J., LI, B., FLORENS, L., SUGANUMA, T., SWANSON, S. K., LEE, K. K., SHIA, W. J., ANDERSON, S., YATES, J., WASHBURN, M. P. & WORKMAN, J. L. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell, 123, 581-92.

CARTER, D., CHAKALOVA, L., OSBORNE, C. S., DAI, Y. F. & FRASER, P. 2002. Long-range chromatin regulatory interactions in vivo. Nat Genet, 32, 623-6.

CHARNEY, R. M., FOROUZMAND, E., CHO, J. S., CHEUNG, J., PARAISO, K. D., YASUOKA, Y., TAKAHASHI, S., TAIRA, M., BLITZ, I. L., XIE, X. & CHO, K. W. 2017. Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program. Dev Cell, 40, 595-607 e4.

CHEN, X., RUBOCK, M. J. & WHITMAN, M. 1996. A transcriptional partner for MAD proteins in TGF-beta signalling. Nature, 383, 691-6.

CHEN, X., WEISBERG, E., FRIDMACHER, V., WATANABE, M., NACO, G. & WHITMAN, M. 1997. Smad4 and FAST-1 in the assembly of activin-responsive factor. Nature, 389, 85-9.

CHEN, X., XU, H., YUAN, P., FANG, F., HUSS, M., VEGA, V. B., WONG, E., ORLOV, Y. L., ZHANG, W., JIANG, J., LOH, Y. H., YEO, H. C., YEO, Z. X., NARANG, V., GOVINDARAJAN, K. R., LEONG, B., SHAHAB, A., RUAN, Y., BOURQUE, G., SUNG, W. K., CLARKE, N. D., WEI, C. L. & NG, H. H. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell, 133, 1106-17.

CHIU, W. T., CHARNEY LE, R., BLITZ, I. L., FISH, M. B., LI, Y., BIESINGER, J., XIE, X. & CHO, K. W. 2014. Genome-wide view of TGFbeta/Foxh1 regulation of the early mesendoderm program. Development, 141, 4537-47.

CHO, H., ORPHANIDES, G., SUN, X., YANG, X. J., OGRYZKO, V., LEES, E., NAKATANI, Y. & REINBERG, D. 1998. A human RNA polymerase II complex containing factors that modify chromatin structure. Mol Cell Biol, 18, 5355-63.

CHO, S., SCHROEDER, S., KAEHLCKE, K., KWON, H. S., PEDAL, A., HERKER, E., SCHNOELZER, M. & OTT, M. 2009. Acetylation of cyclin T1 regulates the equilibrium between active and inactive P-TEFb in cells. EMBO J, 28, 1407-17.

CHUNG, J. H., WHITELEY, M. & FELSENFELD, G. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell, 74, 505-14.

COLLART, C., ALLEN, G. E., BRADSHAW, C. R., SMITH, J. C. & ZEGERMAN, P. 2013. Titration of four replication factors is essential for the Xenopus laevis midblastula transition. Science, 341, 893-6.

COLLART, C., OWENS, N. D., BHAW-ROSUN, L., COOPER, B., DE DOMENICO, E., PATRUSHEV, I., SESAY, A. K., SMITH, J. N., SMITH, J. C. & GILCHRIST, M. J. 2014. High-resolution analysis of gene activity during the Xenopus mid-blastula transition. Development, 141, 1927-39.

CORE, L. J., WATERFALL, J. J. & LIS, J. T. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science, 322, 1845-8.

CORRADIN, O., SAIAKHOVA, A., AKHTAR-ZAIDI, B., MYEROFF, L., WILLIS, J., COWPER-SAL LARI, R., LUPIEN, M., MARKOWITZ, S. & SCACHERI, P. C. 2014. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res, 24, 1-13.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A, 107, 21931-6.

DAI, Z., DAI, X., XIANG, Q. & FENG, J. 2009. Robustness of transcriptional regulatory program influences gene expression variability. BMC Genomics, 10, 573.

DAVIDSON, A. J. & ZON, L. I. 2000. Turning mesoderm into blood: the formation of hematopoietic stem cells during embryogenesis. Curr Top Dev Biol, 50, 45-60.

DAVIE, K., JACOBS, J., ATKINS, M., POTIER, D., CHRISTIAENS, V., HALDER, G. & AERTS, S. 2015. Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. PLoS Genet, 11, e1004994.

DE LA CALLE-MUSTIENES, E., LU, Z., CORTES, M., ANDERSEN, B., MODOLELL, J. & GOMEZ-SKARMETA, J. L. 2003. Xenopus Xlmo4 is a GATA cofactor during ventral mesoderm formation and regulates Ldb1 availability at the dorsal mesoderm and the neural plate. Dev Biol, 264, 564-81.

DE LAAT, W. & GROSVELD, F. 2003. Spatial organization of gene expression: the active chromatin hub. Chromosome Res, 11, 447-59.

DE VILLIERS, J., OLSON, L., BANERJI, J. & SCHAFFNER, W. 1983. Analysis of the transcriptional enhancer effect. Cold Spring Harb Symp Quant Biol, 47 Pt 2, 911-9.

DE WIT, E. & DE LAAT, W. 2012. A decade of 3C technologies: insights into nuclear organization. Genes Dev, 26, 11-24.

DEKKER, J., RIPPE, K., DEKKER, M. & KLECKNER, N. 2002. Capturing chromosome conformation. Science, 295, 1306-11.

DENKER, A. & DE LAAT, W. 2016. The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev, 30, 1357-82.

DILLON, N., TRIMBORN, T., STROUBOULIS, J., FRASER, P. & GROSVELD, F. 1997. The effect of distance on long-range chromatin interactions. Mol Cell, 1, 131-9.

DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature, 485, 376-80.

DONZE, D., ADAMS, C. R., RINE, J. & KAMAKAKA, R. T. 1999. The boundaries of the silenced HMR domain in Saccharomyces cerevisiae. Genes Dev, 13, 698-708.

ECKNER, R., EWEN, M. E., NEWSOME, D., GERDES, M., DECAPRIO, J. A., LAWRENCE, J. B. & LIVINGSTON, D. M. 1994. Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev, 8, 869-84.

EMISON, E. S., MCCALLION, A. S., KASHUK, C. S., BUSH, R. T., GRICE, E., LIN, S., PORTNOY, M. E., CUTLER, D. J., GREEN, E. D. & CHAKRAVARTI, A. 2005. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. Nature, 434, 857-63.

ENCODE 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science, 306, 636-40.

ENCODE 2012. An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 57-74.

ENCODE, BIRNEY, E., STAMATOYANNOPOULOS, J. A., DUTTA, A., GUIGO, R., GINGERAS, T. R., MARGULIES, E. H., WENG, Z., SNYDER, M., DERMITZAKIS, E. T., THURMAN, R. E., KUEHN, M. S., TAYLOR, C. M., NEPH, S., KOCH, C. M., ASTHANA, S., MALHOTRA, A., ADZHUBEI, I., GREENBAUM, J. A., ANDREWS, R. M., FLICEK, P., BOYLE, P. J., CAO, H., CARTER, N. P., CLELLAND, G. K., DAVIS, S., DAY, N., DHAMI, P., DILLON, S. C., DORSCHNER, M. O., FIEGLER, H., GIRESI, P. G., GOLDY, J., HAWRYLYCZ, M., HAYDOCK, A., HUMBERT, R., JAMES, K. D., JOHNSON, B. E., JOHNSON, E. M., FRUM, T. T., ROSENZWEIG, E. R., KARNANI, N., LEE, K., LEFEBVRE, G. C., NAVAS, P. A., NERI, F., PARKER, S. C., SABO, P. J., SANDSTROM, R., SHAFER, A., VETRIE, D., WEAVER, M., WILCOX, S., YU, M., COLLINS, F. S., DEKKER, J., LIEB, J. D., TULLIUS, T. D., CRAWFORD, G. E., SUNYAEV, S., NOBLE, W. S., DUNHAM, I., DENOEUD, F., REYMOND, A., KAPRANOV, P., ROZOWSKY, J., ZHENG, D., CASTELO, R., FRANKISH, A., HARROW, J., GHOSH, S., SANDELIN, A., HOFACKER, I. L., BAERTSCH, R., KEEFE, D., DIKE, S., CHENG, J., HIRSCH, H. A., SEKINGER, E. A., LAGARDE, J., ABRIL, J. F., SHAHAB, A., FLAMM, C., FRIED, C., HACKERMULLER, J., HERTEL, J., LINDEMEYER, M., MISSAL, K., TANZER, A., WASHIETL, S., KORBEL, J., EMANUELSSON, O., PEDERSEN, J. S., HOLROYD, N., TAYLOR, R., SWARBRECK, D., MATTHEWS, N., DICKSON, M. C., THOMAS, D. J., WEIRAUCH, M. T., et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 447, 799-816.

ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R., COYNE, M., KU, M., DURHAM, T., KELLIS, M. & BERNSTEIN, B. E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature, 473, 43-9.

FERRARI, K. J., SCELFO, A., JAMMULA, S., CUOMO, A., BAROZZI, I., STUTZER, A., FISCHLE, W., BONALDI, T. & PASINI, D. 2014. Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol Cell, 53, 49-62.

FISHER, S., GRICE, E. A., VINTON, R. M., BESSLING, S. L., URASAKI, A., KAWAKAMI, K. & MCCALLION, A. S. 2006. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc, 1, 1297-305.

FU, J., YOON, H. G., QIN, J. & WONG, J. 2007. Regulation of P-TEFb elongation complex activity by CDK9 acetylation. Mol Cell Biol, 27, 4641-51.

FUJIMI, T. J., MIKOSHIBA, K. & ARUGA, J. 2006. Xenopus Zic4: conservation and diversification of expression profiles and protein function among the Xenopus Zic family. Dev Dyn, 235, 3379-86.

FUKAYA, T., LIM, B. & LEVINE, M. 2016. Enhancer Control of Transcriptional Bursting. Cell, 166, 358-368.

GAO, F., FOAT, B. C. & BUSSEMAKER, H. J. 2004. Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. BMC Bioinformatics, 5, 31.

GAO, X., SEDGWICK, T., SHI, Y. B. & EVANS, T. 1998. Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation. Mol Cell Biol, 18, 2901-11.

GASZNER, M. & FELSENFELD, G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet, 7, 703-13.

GATES, L. A., SHI, J., ROHIRA, A. D., FENG, Q., ZHU, B., BEDFORD, M. T., SAGUM, C. A., JUNG, S. Y., QIN, J., TSAI, M. J., TSAI, S. Y., LI, W., FOULDS, C. E. & O'MALLEY, B. W. 2017. Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. J Biol Chem, 292, 14456-14472.

GAYTHER, S. A., BATLEY, S. J., LINGER, L., BANNISTER, A., THORPE, K., CHIN, S. F., DAIGO, Y., RUSSELL, P., WILSON, A., SOWTER, H. M., DELHANTY, J. D., PONDER, B. A., KOUZARIDES, T. & CALDAS, C. 2000. Mutations truncating the EP300 acetylase in human cancers. Nat Genet, 24, 300-3.

GENTSCH, G. E., PATRUSHEV, I. & SMITH, J. C. 2015. Genome-wide Snapshot of Chromatin Regulators and States in Xenopus Embryos by ChIP-Seq. J Vis Exp.

GENTSCH, G. E. & SMITH, J. C. 2014. Investigating physical chromatin associations across the Xenopus genome by chromatin immunoprecipitation. Cold Spring Harb Protoc, 2014.

GEORGIOU, G. & VAN HEERINGEN, S. J. 2016. fluff: exploratory analysis and visualization of high-throughput sequencing data. PeerJ, 4, e2209.

GHAVI-HELM, Y., KLEIN, F. A., PAKOZDI, T., CIGLAR, L., NOORDERMEER, D., HUBER, W. & FURLONG, E. E. 2014. Enhancer loops appear stable during development and are associated with paused polymerase. Nature, 512, 96-100.

GILES, R. H., PETERS, D. J. & BREUNING, M. H. 1998. Conjunction dysfunction: CBP/p300 in human disease. Trends Genet, 14, 178-83.

GILMOUR, D. S. & LIS, J. T. 1984. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. Proc Natl Acad Sci U S A, 81, 4275-9.

GITTER, A., SIEGFRIED, Z., KLUTSTEIN, M., FORNES, O., OLIVA, B., SIMON, I. & BAR-JOSEPH, Z. 2009. Backup in gene regulatory networks explains differences between binding and knockout results. Mol Syst Biol, 5, 276.

GOODMAN, R. H. & SMOLIK, S. 2000. CBP/p300 in cell growth, transformation, and development. Genes Dev, 14, 1553-77.

GOVE, C., WALMSLEY, M., NIJJAR, S., BERTWISTLE, D., GUILLE, M., PARTINGTON, G., BOMFORD, A. & PATIENT, R. 1997. Over-expression of GATA-6 in Xenopus embryos blocks differentiation of heart precursors. EMBO J, 16, 355-68.

GUO, Z., ZHENG, L., LIAO, X. & GELLER, D. 2016. Up-Regulation of Human Inducible Nitric Oxide Synthase by p300 Transcriptional Complex. PLoS One, 11, e0146640.

HAMLET, M. R., YERGEAU, D. A., KULIYEV, E., TAKEDA, M., TAIRA, M., KAWAKAMI, K. & MEAD, P. E. 2006. Tol2 transposon-mediated transgenesis in Xenopus tropicalis. Genesis, 44, 438-45.

HARLOW, E., WHYTE, P., FRANZA, B. R., JR. & SCHLEY, C. 1986. Association of adenovirus early-region 1A proteins with cellular polypeptides. Mol Cell Biol, 6, 1579-89.

HAUSEN, P. & RIEBESELL, M. 1991. The Early Development of Xenopus Laevis: An Atlas of the Histology, Germany, Verlag der Zeitschrift für Naturforschung.

HECHT, A., VLEMINCKX, K., STEMMLER, M. P., VAN ROY, F. & KEMLER, R. 2000. The p300/CBP acetyltransferases function as transcriptional coactivators of beta-catenin in vertebrates. EMBO J, 19, 1839-50.

HEINTZMAN, N. D., HON, G. C., HAWKINS, R. D., KHERADPOUR, P., STARK, A., HARP, L. F., YE, Z., LEE, L. K., STUART, R. K., CHING, C. W., CHING, K. A., ANTOSIEWICZ-BOURGET, J. E., LIU, H., ZHANG, X., GREEN, R. D., LOBANENKOV, V. V., STEWART, R., THOMSON, J. A., CRAWFORD, G. E.,

KELLIS, M. & REN, B. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature, 459, 108-12.

HEINTZMAN, N. D., STUART, R. K., HON, G., FU, Y., CHING, C. W., HAWKINS, R. D., BARRERA, L. O., VAN CALCAR, S., QU, C., CHING, K. A., WANG, W., WENG, Z., GREEN, R. D., CRAWFORD, G. E. & REN, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet, 39, 311-8.

HELLSTEN, U., HARLAND, R. M., GILCHRIST, M. J., HENDRIX, D., JURKA, J., KAPITONOV, V., OVCHARENKO, I., PUTNAM, N. H., SHU, S., TAHER, L., BLITZ, I. L., BLUMBERG, B., DICHMANN, D. S., DUBCHAK, I., AMAYA, E., DETTER, J. C., FLETCHER, R., GERHARD, D. S., GOODSTEIN, D., GRAVES, T., GRIGORIEV, I. V., GRIMWOOD, J., KAWASHIMA, T., LINDQUIST, E., LUCAS, S. M., MEAD, P. E., MITROS, T., OGINO, H., OHTA, Y., POLIAKOV, A. V., POLLET, N., ROBERT, J., SALAMOV, A., SATER, A. K., SCHMUTZ, J., TERRY, A., VIZE, P. D., WARREN, W. C., WELLS, D., WILLS, A., WILSON, R. K., ZIMMERMAN, L. B., ZORN, A. M., GRAINGER, R., GRAMMER, T., KHOKHA, M. K., RICHARDSON, P. M. & ROKHSAR, D. S. 2010. The genome of the Western clawed frog Xenopus tropicalis. Science, 328, 633-6.

HILTON, I. B., D'IPPOLITO, A. M., VOCKLEY, C. M., THAKORE, P. I., CRAWFORD, G. E., REDDY, T. E. & GERSBACH, C. A. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol, 33, 510-7.

HOLDRIDGE, C. & DORSETT, D. 1991. Repression of hsp70 heat shock gene transcription by the suppressor of hairy-wing protein of Drosophila melanogaster. Mol Cell Biol, 11, 1894-900.

HOLMQVIST, P. H. & MANNERVIK, M. 2013. Genomic occupancy of the transcriptional co-activators p300 and CBP. Transcription, 4, 18-23.

HONTELEZ, S., VAN KRUIJSBERGEN, I., GEORGIOU, G., VAN HEERINGEN, S. J., BOGDANOVIC, O., LISTER, R. & VEENSTRA, G. J. 2015. Embryonic transcription is controlled by maternally defined chromatin state. Nat Commun, 6, 10148.

HOWE, J. A., MYMRYK, J. S., EGAN, C., BRANTON, P. E. & BAYLEY, S. T. 1990. Retinoblastoma growth suppressor and a 300-kDa protein appear to regulate cellular DNA synthesis. Proc Natl Acad Sci U S A, 87, 5883-7.

HOWELL, M., INMAN, G. J. & HILL, C. S. 2002. A novel Xenopus Smad-interacting forkhead transcription factor (XFast-3) cooperates with XFast-1 in regulating gastrulation movements. Development, 129, 2823-34.

HUNTER, J. D. 2007. Matplotlib: A 2D graphics environment. Computing In Science & Engineering, 9, 90-95.

INOUE, H., TAKAHASHI, H., HASHIMURA, M., ESHIMA, K., AKIYA, M., MATSUMOTO, T. & SAEGUSA, M. 2016. Cooperation of Sox4 with beta-catenin/p300 complex in transcriptional regulation of the Slug gene during divergent sarcomatous differentiation in uterine carcinosarcoma. BMC Cancer, 16, 53.

INOUE, Y., ITOH, Y., ABE, K., OKAMOTO, T., DAITOKU, H., FUKAMIZU, A., ONOZAKI, K. & HAYASHI, H. 2007. Smad3 is acetylated by p300/CBP to regulate its transactivation activity. Oncogene, 26, 500-8.

IOTT, N. E., HEWARD, J. A., ROUX, B., TSITSIOU, E., FENWICK, P. S., LENZI, L., GOODHEAD, I., HERTZ-FOWLER, C., HEGER, A., HALL, N., DONNELLY, L. E., SIMS, D. & LINDSAY, M. A. 2014. Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. Nat Commun, 5, 3979.

IYER, N. G., OZDAG, H. & CALDAS, C. 2004. p300/CBP and cancer. Oncogene, 23, 4225-31.

IYER, V. R., HORAK, C. E., SCAFE, C. S., BOTSTEIN, D., SNYDER, M. & BROWN, P. O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature, 409, 533-8.

JANKNECHT, R. & HUNTER, T. 1996a. Transcription. A growing coactivator network. Nature, 383, 22-3.

JANKNECHT, R. & HUNTER, T. 1996b. Versatile molecular glue. Transcriptional control. Curr Biol, 6, 951-4.

JAVIERRE, B. M., BURREN, O. S., WILDER, S. P., KREUZHUBER, R., HILL, S. M., SEWITZ, S., CAIRNS, J., WINGETT, S. W., VARNAI, C., THIECKE, M. J., BURDEN, F., FARROW, S., CUTLER, A. J., REHNSTROM, K., DOWNES, K., GRASSI, L., KOSTADIMA, M., FREIRE-PRITCHETT, P., WANG, F., CONSORTIUM, B., STUNNENBERG, H. G., TODD, J. A., ZERBINO, D. R., STEGLE, O., OUWEHAND, W. H., FRONTINI, M., WALLACE, C., SPIVAKOV, M. & FRASER, P. 2016. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell, 167, 1369-1384 e19.

JI, X., DADON, D. B., POWELL, B. E., FAN, Z. P., BORGES-RIVERA, D., SHACHAR, S., WEINTRAUB, A. S., HNISZ, D., PEGORARO, G., LEE, T. I., MISTELI, T., JAENISCH, R. & YOUNG, R. A. 2016. 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell, 18, 262-75.

JIANG, Y. & EVANS, T. 1996. The Xenopus GATA-4/5/6 genes are associated with cardiac specification and can regulate cardiac-specific transcription during embryogenesis. Dev Biol, 174, 258-70.

JIN, Q., YU, L. R., WANG, L., ZHANG, Z., KASPER, L. H., LEE, J. E., WANG, C., BRINDLE, P. K., DENT, S. Y. & GE, K. 2011. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J, 30, 249-62.

JOHNSON, D. S., MORTAZAVI, A., MYERS, R. M. & WOLD, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science, 316, 1497-502.

JONKERS, I., KWAK, H. & LIS, J. T. 2014. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife, 3, e02407.

JORSTAD, N. L., WILKEN, M. S., GRIMES, W. N., WOHL, S. G., VANDENBOSCH, L. S., YOSHIMATSU, T., WONG, R. O., RIEKE, F. & REH, T. A. 2017. Stimulation of functional neuronal regeneration from Muller glia in adult mice. Nature, 548, 103-107.

JOSHI, A. A. & STRUHL, K. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Mol Cell, 20, 971-8.

KADONAGA, J. T. 2004. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. Cell, 116, 247-57.

KAGEY, M. H., NEWMAN, J. J., BILODEAU, S., ZHAN, Y., ORLANDO, D. A., VAN BERKUM, N. L., EBMEIER, C. C., GOOSSENS, J., RAHL, P. B., LEVINE, S. S., TAATJES, D. J., DEKKER, J. & YOUNG, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. Nature, 467, 430-5.

KARMODIYA, K., KREBS, A. R., OULAD-ABDELGHANI, M., KIMURA, H. & TORA, L. 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics, 13, 424.

KARPINKA, J. B., FORTRIEDE, J. D., BURNS, K. A., JAMES-ZORN, C., PONFERRADA, V. G., LEE, J., KARIMI, K., ZORN, A. M. & VIZE, P. D. 2015.

Xenbase, the Xenopus model organism database; new virtualized system, data types and genomes. Nucleic Acids Res, 43, D756-63.

KASPER, L. H., QU, C., OBENAUER, J. C., MCGOLDRICK, D. J. & BRINDLE, P. K. 2014. Genome-wide and single-cell analyses reveal a context dependent relationship between CBP recruitment and gene expression. Nucleic Acids Res, 42, 11363-82.

KAWAKAMI, K. 2007. Tol2: a versatile gene transfer vector in vertebrates. Genome Biol, 8 Suppl 1, S7.

KAWAKAMI, K., IMANAKA, K., ITOH, M. & TAIRA, M. 2004. Excision of the Tol2 transposable element of the medaka fish Oryzias latipes in Xenopus laevis and Xenopus tropicalis. Gene, 338, 93-8.

KEE, B. L., ARIAS, J. & MONTMINY, M. R. 1996. Adaptor-mediated recruitment of RNA polymerase II to a signal-dependent activator. J Biol Chem, 271, 2373-5.

KELLUM, R. & SCHEDL, P. 1991. A position-effect assay for boundaries of higher order chromosomal domains. Cell, 64, 941-50.

KELLUM, R. & SCHEDL, P. 1992. A group of scs elements function as domain boundaries in an enhancer-blocking assay. Mol Cell Biol, 12, 2424-31.

KEOGH, M. C., KURDISTANI, S. K., MORRIS, S. A., AHN, S. H., PODOLNY, V., COLLINS, S. R., SCHULDINER, M., CHIN, K., PUNNA, T., THOMPSON, N. J., BOONE, C., EMILI, A., WEISSMAN, J. S., HUGHES, T. R., STRAHL, B. D., GRUNSTEIN, M., GREENBLATT, J. F., BURATOWSKI, S. & KROGAN, N. J. 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell, 123, 593-605.

KIM, T. K., HEMBERG, M., GRAY, J. M., COSTA, A. M., BEAR, D. M., WU, J., HARMIN, D. A., LAPTEWICZ, M., BARBARA-HALEY, K., KUERSTEN, S., MARKENSCOFF-PAPADIMITRIOU, E., KUHL, D., BITO, H., WORLEY, P. F., KREIMAN, G. & GREENBERG, M. E. 2010. Widespread transcription at neuronal activity-regulated enhancers. Nature, 465, 182-7.

KIM, Y. W. & KIM, A. 2013. Histone acetylation contributes to chromatin looping between the locus control region and globin gene by influencing hypersensitive site formation. Biochim Biophys Acta, 1829, 963-9.

KITAGUCHI, T., NAGAI, T., NAKATA, K., ARUGA, J. & MIKOSHIBA, K. 2000. Zic3 is involved in the left-right specification of the Xenopus embryo. Development, 127, 4787-95.

KIYOTA, T., KATO, A., ALTMANN, C. R. & KATO, Y. 2008. The POU homeobox protein Oct-1 regulates radial glia formation downstream of Notch signaling. Dev Biol, 315, 579-92.

KOCH, C. M., ANDREWS, R. M., FLICEK, P., DILLON, S. C., KARAOZ, U., CLELLAND, G. K., WILCOX, S., BEARE, D. M., FOWLER, J. C., COUTTET, P., JAMES, K. D., LEFEBVRE, G. C., BRUCE, A. W., DOVEY, O. M., ELLIS, P. D., DHAMI, P., LANGFORD, C. F., WENG, Z., BIRNEY, E., CARTER, N. P., VETRIE, D. & DUNHAM, I. 2007. The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res, 17, 691-707.

KOUZARIDES, T. 2007. Chromatin modifications and their function. Cell, 128, 693-705.

KREJCI, A., BERNARD, F., HOUSDEN, B. E., COLLINS, S. & BRAY, S. J. 2009. Direct response to Notch activation: signaling crosstalk and incoherent logic. Sci Signal, 2, ra1.

KROGAN, N. J., KIM, M., TONG, A., GOLSHANI, A., CAGNEY, G., CANADIEN, V., RICHARDS, D. P., BEATTIE, B. K., EMILI, A., BOONE, C., SHILATIFARD, A., BURATOWSKI, S. & GREENBLATT, J. 2003. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol, 23, 4207-18.

KUO, C. T., MORRISEY, E. E., ANANDAPPA, R., SIGRIST, K., LU, M. M., PARMACEK, M. S., SOUDAIS, C. & LEIDEN, J. M. 1997. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes Dev, 11, 1048-60.

KUO, J. S., PATEL, M., GAMSE, J., MERZDORF, C., LIU, X., APEKIN, V. & SIVE, H. 1998. Opl: a zinc finger protein that regulates neural determination and patterning in Xenopus. Development, 125, 2867-82.

KUZMICHEV, A., NISHIOKA, K., ERDJUMENT-BROMAGE, H., TEMPST, P. & REINBERG, D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev, 16, 2893-905.

LACHNER, M., O'CARROLL, D., REA, S., MECHTLER, K. & JENUWEIN, T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature, 410, 116-20.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods, 9, 357-9.

LEE, S. K., ANZICK, S. L., CHOI, J. E., BUBENDORF, L., GUAN, X. Y., JUNG, Y. K., KALLIONIEMI, O. P., KONONEN, J., TRENT, J. M., AZORSA, D., JHUN, B. H., CHEONG, J. H., LEE, Y. C., MELTZER, P. S. & LEE, J. W. 1999. A nuclear factor, ASC-2, as a cancer-amplified transcriptional coactivator essential for ligand-dependent transactivation by nuclear receptors in vivo. J Biol Chem, 274, 34283-93.

LETTICE, L. A., HEANEY, S. J., PURDIE, L. A., LI, L., DE BEER, P., OOSTRA, B. A., GOODE, D., ELGAR, G., HILL, R. E. & DE GRAAFF, E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet, 12, 1725-35.

LEVINE, M., CATTOGLIO, C. & TJIAN, R. 2014. Looping back to leap forward: transcription enters a new era. Cell, 157, 13-25.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25, 2078-9.

LI, W., NOTANI, D., MA, Q., TANASA, B., NUNEZ, E., CHEN, A. Y., MERKURJEV, D., ZHANG, J., OHGI, K., SONG, X., OH, S., KIM, H. S., GLASS, C. K. & ROSENFELD, M. G. 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature, 498, 516-20.

LIEB, J. D., LIU, X., BOTSTEIN, D. & BROWN, P. O. 2001. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet, 28, 327-34.

LIN, C., GARRUSS, A. S., LUO, Z., GUO, F. & SHILATIFARD, A. 2013. The RNA Pol II elongation factor Ell3 marks enhancers in ES cells and primes future gene activation. Cell, 152, 144-56.

LIN, Y. C., BENNER, C., MANSSON, R., HEINZ, S., MIYAZAKI, K., MIYAZAKI, M., CHANDRA, V., BOSSEN, C., GLASS, C. K. & MURRE, C. 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat Immunol, 13, 1196-204.

LIU, F., POUPONNOT, C. & MASSAGUE, J. 1997. Dual role of the Smad4/DPC4 tumor suppressor in TGFbeta-inducible transcriptional complexes. Genes Dev, 11, 3157-67.

LIU, F., VAN DEN BROEK, O., DESTREE, O. & HOPPLER, S. 2005. Distinct roles for Xenopus Tcf/Lef genes in mediating specific responses to Wnt/beta-catenin signalling in mesoderm development. Development, 132, 5375-85.

LIU, L., SCOLNICK, D. M., TRIEVEL, R. C., ZHANG, H. B., MARMORSTEIN, R., HALAZONETIS, T. D. & BERGER, S. L. 1999. p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage. Mol Cell Biol, 19, 1202-9.

LOOTS, G. G., BERGMANN, A., HUM, N. R., OLDENBURG, C. E., WILLS, A. E., HU, N., OVCHARENKO, I. & HARLAND, R. M. 2013. Interrogating transcriptional regulatory sequences in Tol2-mediated Xenopus transgenics. PLoS One, 8, e68548.

LU, F., LIU, Y., INOUE, A., SUZUKI, T., ZHAO, K. & ZHANG, Y. 2016. Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. Cell, 165, 1375-1388.

LUNDBLAD, J. R., KWOK, R. P., LAURANCE, M. E., HARTER, M. L. & GOODMAN, R. H. 1995. Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP. Nature, 374, 85-8.

LUPIANEZ, D. G., KRAFT, K., HEINRICH, V., KRAWITZ, P., BRANCATI, F., KLOPOCKI, E., HORN, D., KAYSERILI, H., OPITZ, J. M., LAXOVA, R., SANTOS-SIMARRO, F., GILBERT-DUSSARDIER, B., WITTLER, L., BORSCHIWER, M., HAAS, S. A., OSTERWALDER, M., FRANKE, M., TIMMERMANN, B., HECHT, J., SPIELMANN, M., VISEL, A. & MUNDLOS, S. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell, 161, 1012-1025.

MA, H., NGUYEN, C., LEE, K. S. & KAHN, M. 2005. Differential roles for the coactivators CBP and p300 on TCF/beta-catenin-mediated survivin gene expression. Oncogene, 24, 3619-31.

MADSEN, J. G., SCHMIDT, S. F., LARSEN, B. D., LOFT, A., NIELSEN, R. & MANDRUP, S. 2015. iRNA-seq: computational method for genome-wide assessment of acute transcriptional regulation from total RNA-seq data. Nucleic Acids Res, 43, e40.

MASTON, G. A., EVANS, S. K. & GREEN, M. R. 2006. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet, 7, 29-59.

MAY, D., BLOW, M. J., KAPLAN, T., MCCULLEY, D. J., JENSEN, B. C., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., AFZAL, V., SIMPSON, P. C., RUBIN, E. M., BLACK, B. L., BRISTOW, J., PENNACCHIO, L. A. & VISEL, A. 2011. Large-scale discovery of enhancers from human heart tissue. Nat Genet, 44, 89-93.

MCGAUGHEY, D. M., VINTON, R. M., HUYNH, J., AL-SAIF, A., BEER, M. A. & MCCALLION, A. S. 2008. Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at phox2b. Genome Res, 18, 252-60.

MEAD, P. E., DECONINCK, A. E., HUBER, T. L., ORKIN, S. H. & ZON, L. I. 2001. Primitive erythropoiesis in the Xenopus embryo: the synergistic role of LMO-2, SCL and GATA-binding proteins. Development, 128, 2301-8.

MESSENGER, N. J. & WARNER, A. E. 1989. The appearance of neural and glial cell markers during early development of the nervous system in the amphibian embryo. Development, 107, 43-54.

MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T. K., KOCHE, R. P., LEE, W., MENDENHALL, E., O'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. & BERNSTEIN, B. E. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature, 448, 553-60.

MORRISON, G. M. & BRICKMAN, J. M. 2006. Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. Development, 133, 2011-22.

MOUSAVI, K., ZARE, H., DELL'ORSO, S., GRONTVED, L., GUTIERREZ-CRUZ, G., DERFOUL, A., HAGER, G. L. & SARTORELLI, V. 2013. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell, 51, 606-17.

MULLEN, A. C., ORLANDO, D. A., NEWMAN, J. J., LOVEN, J., KUMAR, R. M., BILODEAU, S., REDDY, J., GUENTHER, M. G., DEKOTER, R. P. & YOUNG, R. A. 2011. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell, 147, 565-76.

NAKAMURA, Y., DE PAIVA ALVES, E., VEENSTRA, G. J. & HOPPLER, S. 2016. Tissue- and stage-specific Wnt target gene expression is controlled subsequent to beta-catenin recruitment to cis-regulatory modules. Development, 143, 1914-25.

NAKATA, K., NAGAI, T., ARUGA, J. & MIKOSHIBA, K. 1997. Xenopus Zic3, a primary regulator both in neural and neural crest development. Proc Natl Acad Sci U S A, 94, 11980-5.

NAKATA, K., NAGAI, T., ARUGA, J. & MIKOSHIBA, K. 1998. Xenopus Zic family and its role in neural and neural crest development. Mech Dev, 75, 43-51.

NAKAYAMA, J., RICE, J. C., STRAHL, B. D., ALLIS, C. D. & GREWAL, S. I. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science, 292, 110-3.

NAM, J. & DAVIDSON, E. H. 2012. Barcoded DNA-tag reporters for multiplex cis-regulatory analysis. PLoS One, 7, e35934.

NEMER, G., QURESHI, S. T., MALO, D. & NEMER, M. 1999. Functional analysis and chromosomal mapping of Gata5, a gene encoding a zinc finger DNA-binding protein. Mamm Genome, 10, 993-9.

NEWPORT, J. & KIRSCHNER, M. 1982a. A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell, 30, 675-86.

NEWPORT, J. & KIRSCHNER, M. 1982b. A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. Cell, 30, 687-96.

NIEUWKOOP, P. D. & FABER, J. 1994. Normal Table of Xenopus Laevis (Daudin): A Systematical & Chronological Survey of the Development from the Fertilized Egg till the End of Metamorphosis, Garland Science.

NISHIOKA, K., CHUIKOV, S., SARMA, K., ERDJUMENT-BROMAGE, H., ALLIS, C. D., TEMPST, P. & REINBERG, D. 2002a. Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev, 16, 479-89.

NISHIOKA, K., RICE, J. C., SARMA, K., ERDJUMENT-BROMAGE, H., WERNER, J., WANG, Y., CHUIKOV, S., VALENZUELA, P., TEMPST, P., STEWARD, R., LIS, J. T., ALLIS, C. D. & REINBERG, D. 2002b. PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. Mol Cell, 9, 1201-13.

NISHITANI, E., LI, C., LEE, J., HOTTA, H., KATAYAMA, Y., YAMAGUCHI, M. & KINOSHITA, T. 2015. Pou5f3.2-induced proliferative state of embryonic cells during gastrulation of Xenopus laevis embryo. Dev Growth Differ, 57, 591-600.

NOMA, K., ALLIS, C. D. & GREWAL, S. I. 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science, 293, 1150-5.

NORA, E. P., LAJOIE, B. R., SCHULZ, E. G., GIORGETTI, L., OKAMOTO, I., SERVANT, N., PIOLOT, T., VAN BERKUM, N. L., MEISIG, J., SEDAT, J., GRIBNAU, J., BARILLOT, E., BLUTHGEN, N., DEKKER, J. & HEARD, E. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature, 485, 381-5.

NOWLING, T., BERNADT, C., JOHNSON, L., DESLER, M. & RIZZINO, A. 2003. The co-activator p300 associates physically with and can mediate the action of the distal enhancer of the FGF-4 gene. J Biol Chem, 278, 13696-705.

O'KANE, C. J. & GEHRING, W. J. 1987. Detection in situ of genomic regulatory elements in Drosophila. Proc Natl Acad Sci U S A, 84, 9123-7.

ODOM, D. T., DOWELL, R. D., JACOBSEN, E. S., GORDON, W., DANFORD, T. W., MACISAAC, K. D., ROLFE, P. A., CONBOY, C. M., GIFFORD, D. K. & FRAENKEL, E. 2007. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat Genet, 39, 730-2.

OIKE, Y., HATA, A., MAMIYA, T., KANAME, T., NODA, Y., SUZUKI, M., YASUE, H., NABESHIMA, T., ARAKI, K. & YAMAMURA, K. 1999a. Truncated CBP protein leads to classical Rubinstein-Taybi syndrome phenotypes in mice: implications for a dominant-negative mechanism. Hum Mol Genet, 8, 387-96.

OIKE, Y., TAKAKURA, N., HATA, A., KANAME, T., AKIZUKI, M., YAMAGUCHI, Y., YASUE, H., ARAKI, K., YAMAMURA, K. & SUDA, T. 1999b. Mice homozygous for a truncated form of CREB-binding protein exhibit defects in hematopoiesis and vasculo-angiogenesis. Blood, 93, 2771-9.

OWENS, N. D., BLITZ, I. L., LANE, M. A., PATRUSHEV, I., OVERTON, J. D., GILCHRIST, M. J., CHO, K. W. & KHOKHA, M. K. 2016. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. Cell Rep, 14, 632-47.

PALSTRA, R. J., TOLHUIS, B., SPLINTER, E., NIJMEIJER, R., GROSVELD, F. & DE LAAT, W. 2003. The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet, 35, 190-4.

PATWARDHAN, R. P., HIATT, J. B., WITTEN, D. M., KIM, M. J., SMITH, R. P., MAY, D., LEE, C., ANDRIE, J. M., LEE, S. I., COOPER, G. M., AHITUV, N., PENNACCHIO, L. A. & SHENDURE, J. 2012. Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol, 30, 265-70.

PETERS, A. H., MERMOUD, J. E., O'CARROLL, D., PAGANI, M., SCHWEIZER, D., BROCKDORFF, N. & JENUWEIN, T. 2002. Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nat Genet, 30, 77-80.

PHILLIPS-CREMINS, J. E., SAURIA, M. E., SANYAL, A., GERASIMOVA, T. I., LAJOIE, B. R., BELL, J. S., ONG, C. T., HOOKWAY, T. A., GUO, C., SUN, Y., BLAND, M. J., WAGSTAFF, W., DALTON, S., MCDEVITT, T. C., SEN, R., DEKKER, J., TAYLOR, J. & CORCES, V. G. 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell, 153, 1281-95.

PIFER, P. M., FARRIS, J. C., THOMAS, A. L., STOILOV, P., DENVIR, J., SMITH, D. M. & FRISCH, S. M. 2016. Grainyhead-like 2 inhibits the coactivator p300, suppressing tubulogenesis and the epithelial-mesenchymal transition. Mol Biol Cell, 27, 2479-92.

PLATH, K., FANG, J., MLYNARCZYK-EVANS, S. K., CAO, R., WORRINGER, K. A., WANG, H., DE LA CRUZ, C. C., OTTE, A. P., PANNING, B. & ZHANG, Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. Science, 300, 131-5.

PRIOLEAU, M. N., NONY, P., SIMPSON, M. & FELSENFELD, G. 1999. An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene. EMBO J, 18, 4035-48.

PTASHNE, M. 1986. Gene regulation by proteins acting nearby and at a distance. Nature, 322, 697-701.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26, 841-2.

RADA-IGLESIAS, A., BAJPAI, R., SWIGUT, T., BRUGMANN, S. A., FLYNN, R. A. & WYSOCKA, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. Nature, 470, 279-83.

RAMIREZ, F., RYAN, D. P., GRUNING, B., BHARDWAJ, V., KILPERT, F., RICHTER, A. S., HEYNE, S., DUNDAR, F. & MANKE, T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res, 44, W160-5.

READ, E. M., RODAWAY, A. R., NEAVE, B., BRANDON, N., HOLDER, N., PATIENT, R. K. & WALMSLEY, M. E. 1998. Evidence for non-axial A/P patterning in the nonneural ectoderm of Xenopus and zebrafish pregastrula embryos. Int J Dev Biol, 42, 763-74.

REN, B., ROBERT, F., WYRICK, J. J., APARICIO, O., JENNINGS, E. G., SIMON, I., ZEITLINGER, J., SCHREIBER, J., HANNETT, N., KANIN, E., VOLKERT, T. L., WILSON, C. J., BELL, S. P. & YOUNG, R. A. 2000. Genome-wide location and function of DNA binding proteins. Science, 290, 2306-9.

RHEE, H. S. & PUGH, B. F. 2012. ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. Curr Protoc Mol Biol, Chapter 21, Unit 21 24.

RICKELS, R., HERZ, H. M., SZE, C. C., CAO, K., MORGAN, M. A., COLLINGS, C. K., GAUSE, M., TAKAHASHI, Y. H., WANG, L., RENDLEMAN, E. J., MARSHALL, S. A., KRUEGER, A., BARTOM, E. T., PIUNTI, A., SMITH, E. R., ABSHIRU, N. A., KELLEHER, N. L., DORSETT, D. & SHILATIFARD, A. 2017. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat Genet, 49, 1647-1653.

ROBERTS, T. C., HART, J. R., KAIKKONEN, M. U., WEINBERG, M. S., VOGT, P. K. & MORRIS, K. V. 2015. Quantification of nascent transcription by bromouridine immunocapture nuclear run-on RT-qPCR. Nat Protoc, 10, 1198-211.

ROBERTSON, G., HIRST, M., BAINBRIDGE, M., BILENKY, M., ZHAO, Y., ZENG, T., EUSKIRCHEN, G., BERNIER, B., VARHOL, R., DELANEY, A., THIESSEN, N., GRIFFITH, O. L., HE, A., MARRA, M., SNYDER, M. & JONES, S. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods, 4, 651-7.

ROEL, G., GENT, Y. Y., PETERSON-MADURO, J., VERBEEK, F. J. & DESTREE, O. 2009. Lef1 plays a role in patterning the mesoderm and ectoderm in Xenopus tropicalis. Int J Dev Biol, 53, 81-9.

ROEL, G., HAMILTON, F. S., GENT, Y., BAIN, A. A., DESTREE, O. & HOPPLER, S. 2002. Lef-1 and Tcf-3 transcription factors mediate tissue-specific Wnt signaling during Xenopus development. Curr Biol, 12, 1941-5.

ROELFSEMA, J. H., WHITE, S. J., ARIYUREK, Y., BARTHOLDI, D., NIEDRIST, D., PAPADIA, F., BACINO, C. A., DEN DUNNEN, J. T., VAN OMMEN, G. J., BREUNING, M. H., HENNEKAM, R. C. & PETERS, D. J. 2005. Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. Am J Hum Genet, 76, 572-80.

ROGERS, C. D., HARAFUJI, N., ARCHER, T., CUNNINGHAM, D. D. & CASEY, E. S. 2009. Xenopus Sox3 activates sox2 and geminin and indirectly represses Xvent2 expression to induce neural progenitor formation at the expense of non-neural ectodermal derivatives. Mech Dev, 126, 42-55.

ROH, T. Y., CUDDAPAH, S. & ZHAO, K. 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev, 19, 542-52.

ROH, T. Y., WEI, G., FARRELL, C. M. & ZHAO, K. 2007. Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Res, 17, 74-81.

ROSS, S., CHEUNG, E., PETRAKIS, T. G., HOWELL, M., KRAUS, W. L. & HILL, C. S. 2006. Smads orchestrate specific histone modifications and chromatin remodeling to activate transcription. EMBO J, 25, 4490-502.

ROTH, S. Y., DENU, J. M. & ALLIS, C. D. 2001. Histone acetyltransferases. Annu Rev Biochem, 70, 81-120.

RUF, S., SYMMONS, O., USLU, V. V., DOLLE, D., HOT, C., ETTWILLER, L. & SPITZ, F. 2011. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat Genet, 43, 379-86.

SANTER, F. R., HOSCHELE, P. P., OH, S. J., ERB, H. H., BOUCHAL, J., CAVARRETTA, I. T., PARSON, W., MEYERS, D. J., COLE, P. A. & CULIG, Z. 2011. Inhibition of the acetyltransferases p300 and CBP reveals a targetable function for p300 in the survival and invasion pathways of prostate cancer cell lines. Mol Cancer Ther, 10, 1644-55.

SANTOS-ROSA, H., SCHNEIDER, R., BANNISTER, A. J., SHERRIFF, J., BERNSTEIN, B. E., EMRE, N. C., SCHREIBER, S. L., MELLOR, J. & KOUZARIDES, T. 2002. Active genes are tri-methylated at K4 of histone H3. Nature, 419, 407-11.

SCHAUKOWITCH, K., JOO, J. Y., LIU, X., WATTS, J. K., MARTINEZ, C. & KIM, T. K. 2014. Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell, 56, 29-42.

SCHMIDT, D., WILSON, M. D., BALLESTER, B., SCHWALIE, P. C., BROWN, G. D., MARSHALL, A., KUTTER, C., WATT, S., MARTINEZ-JIMENEZ, C. P., MACKAY, S., TALIANIDIS, I., FLICEK, P. & ODOM, D. T. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science, 328, 1036-40.

SCHNEIDER, R., BANNISTER, A. J., MYERS, F. A., THORNE, A. W., CRANE-ROBINSON, C. & KOUZARIDES, T. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol, 6, 73-7.

SCHOTTA, G., LACHNER, M., SARMA, K., EBERT, A., SENGUPTA, R., REUTER, G., REINBERG, D. & JENUWEIN, T. 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev, 18, 1251-62.

SCHRODER, S., HERKER, E., ITZEN, F., HE, D., THOMAS, S., GILCHRIST, D. A., KAEHLCKE, K., CHO, S., POLLARD, K. S., CAPRA, J. A., SCHNOLZER, M., COLE, P. A., GEYER, M., BRUNEAU, B. G., ADELMAN, K. & OTT, M. 2013. Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. Mol Cell, 52, 314-24.

SEXTON, T., YAFFE, E., KENIGSBERG, E., BANTIGNIES, F., LEBLANC, B., HOICHMAN, M., PARRINELLO, H., TANAY, A. & CAVALLI, G. 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell, 148, 458-72.

SHI, J., WHYTE, W. A., ZEPEDA-MENDOZA, C. J., MILAZZO, J. P., SHEN, C., ROE, J. S., MINDER, J. L., MERCAN, F., WANG, E., ECKERSLEY-MASLIN, M. A., CAMPBELL, A. E., KAWAOKA, S., SHAREEF, S., ZHU, Z., KENDALL, J., MUHAR, M., HASLINGER, C., YU, M., ROEDER, R. G., WIGLER, M. H., BLOBEL, G. A., ZUBER, J., SPECTOR, D. L., YOUNG, R. A. & VAKOC, C. R. 2013. Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. Genes Dev, 27, 2648-62.

SHIKAMA, N., LEE, C. W., FRANCE, S., DELAVAINE, L., LYON, J., KRSTIC-DEMONACOS, M. & LA THANGUE, N. B. 1999. A novel cofactor for p300 that regulates the p53 response. Mol Cell, 4, 365-76.

SHLYUEVA, D., STAMPFEL, G. & STARK, A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet, 15, 272-86.

SHOWELL, C. & CONLON, F. L. 2009. The western clawed frog (Xenopus tropicalis): an emerging vertebrate model for developmental genetics and environmental toxicology. Cold Spring Harb Protoc, 2009, pdb emo131.

SILVA, J., MAK, W., ZVETKOVA, I., APPANAH, R., NESTEROVA, T. B., WEBSTER, Z., PETERS, A. H., JENUWEIN, T., OTTE, A. P. & BROCKDORFF, N. 2003. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell, 4, 481-95.

SIMONIS, M., KLOUS, P., SPLINTER, E., MOSHKIN, Y., WILLEMSEN, R., DE WIT, E., VAN STEENSEL, B. & DE LAAT, W. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet, 38, 1348-54.

SKIRKANICH, J., LUXARDI, G., YANG, J., KODJABACHIAN, L. & KLEIN, P. S. 2011. An essential role for transcription before the MBT in Xenopus laevis. Dev Biol, 357, 478-91.

SMITS, A. H., LINDEBOOM, R. G., PERINO, M., VAN HEERINGEN, S. J., VEENSTRA, G. J. & VERMEULEN, M. 2014. Global absolute quantification reveals tight regulation of protein expression in single Xenopus eggs. Nucleic Acids Res, 42, 9880-91.

SOLOMON, M. J., LARSEN, P. L. & VARSHAVSKY, A. 1988. Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. Cell, 53, 937-47.

SONG, L. & CRAWFORD, G. E. 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc, 2010, pdb prot5384.

SPIEGELMAN, B. M. & HEINRICH, R. 2004. Biological control through regulated transcriptional coactivators. Cell, 119, 157-67.

SPITZ, F. & FURLONG, E. E. 2012. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet, 13, 613-26.

SPITZ, F., GONZALEZ, F. & DUBOULE, D. 2003. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell, 113, 405-17.

SPLINTER, E., HEATH, H., KOOREN, J., PALSTRA, R. J., KLOUS, P., GROSVELD, F., GALJART, N. & DE LAAT, W. 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev, 20, 2349-54.

STRAHL, B. D., GRANT, P. A., BRIGGS, S. D., SUN, Z. W., BONE, J. R., CALDWELL, J. A., MOLLAH, S., COOK, R. G., SHABANOWITZ, J., HUNT, D. F. & ALLIS, C. D. 2002. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol Cell Biol, 22, 1298-306.

SUN, L., BERTKE, M. M., CHAMPION, M. M., ZHU, G., HUBER, P. W. & DOVICHI, N. J. 2014. Quantitative proteomics of Xenopus laevis embryos: expression kinetics of nearly 4000 proteins during early development. Sci Rep, 4, 4365.

SUN, L., DUBIAK, K. M., PEUCHEN, E. H., ZHANG, Z., ZHU, G., HUBER, P. W. & DOVICHI, N. J. 2016. Single Cell Proteomics Using Frog (Xenopus laevis) Blastomeres Isolated from Early Stage Embryos, Which Form a Geometric Progression in Protein Content. Anal Chem, 88, 6653-7.

SUN, Y., KOLLIGS, F. T., HOTTIGER, M. O., MOSAVIN, R., FEARON, E. R. & NABEL, G. J. 2000. Regulation of beta -catenin transformation by the p300 transcriptional coactivator. Proc Natl Acad Sci U S A, 97, 12613-8.

SUR, I. K., HALLIKAS, O., VAHARAUTIO, A., YAN, J., TURUNEN, M., ENGE, M., TAIPALE, M., KARHU, A., AALTONEN, L. A. & TAIPALE, J. 2012. Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. Science, 338, 1360-3.

TACHIBANA, M., SUGIMOTO, K., NOZAKI, M., UEDA, J., OHTA, T., OHKI, M., FUKUDA, M., TAKEDA, N., NIIDA, H., KATO, H. & SHINKAI, Y. 2002. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev, 16, 1779-91.

TADROS, W. & LIPSHITZ, H. D. 2009. The maternal-to-zygotic transition: a play in two acts. Development, 136, 3033-42.

TAKEMARU, K. I. & MOON, R. T. 2000. The transcriptional coactivator CBP interacts with beta-catenin to activate gene expression. J Cell Biol, 149, 249-54.

TALASZ, H., LINDNER, H. H., SARG, B. & HELLIGER, W. 2005. Histone H4-lysine 20 monomethylation is increased in promoter and coding regions of active genes and correlates with hyperacetylation. J Biol Chem, 280, 38814-22.

TAN, M. H., AU, K. F., YABLONOVITCH, A. L., WILLS, A. E., CHUANG, J., BAKER, J. C., WONG, W. H. & LI, J. B. 2013. RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. Genome Res, 23, 201-16.

TAO, J., KULIYEV, E., WANG, X., LI, X., WILANOWSKI, T., JANE, S. M., MEAD, P. E. & CUNNINGHAM, J. M. 2005. BMP4-dependent expression of Xenopus Grainyhead-like 1 is essential for epidermal differentiation. Development, 132, 1021-34.

THURMAN, R. E., RYNES, E., HUMBERT, R., VIERSTRA, J., MAURANO, M. T., HAUGEN, E., SHEFFIELD, N. C., STERGACHIS, A. B., WANG, H., VERNOT, B., GARG, K., JOHN, S., SANDSTROM, R., BATES, D., BOATMAN, L., CANFIELD, T. K., DIEGEL, M., DUNN, D., EBERSOL, A. K., FRUM, T., GISTE, E., JOHNSON, A. K., JOHNSON, E. M., KUTYAVIN, T., LAJOIE, B., LEE, B. K., LEE, K., LONDON, D., LOTAKIS, D., NEPH, S., NERI, F., NGUYEN, E. D., QU, H., REYNOLDS, A. P., ROACH, V., SAFI, A., SANCHEZ, M. E., SANYAL, A., SHAFER, A., SIMON, J. M., SONG, L., VONG, S., WEAVER, M., YAN, Y., ZHANG, Z., ZHANG, Z., LENHARD, B., TEWARI, M., DORSCHNER, M. O., HANSEN, R. S., NAVAS, P. A., STAMATOYANNOPOULOS, G., IYER, V. R., LIEB, J. D., SUNYAEV, S. R., AKEY, J. M., SABO, P. J., KAUL, R., FUREY, T. S., DEKKER, J., CRAWFORD, G. E. & STAMATOYANNOPOULOS, J. A. 2012. The accessible chromatin landscape of the human genome. Nature, 489, 75-82.

TING, C. N., OLSON, M. C., BARTON, K. P. & LEIDEN, J. M. 1996. Transcription factor GATA-3 is required for development of the T-cell lineage. Nature, 384, 474-8.

TOLHUIS, B., PALSTRA, R. J., SPLINTER, E., GROSVELD, F. & DE LAAT, W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell, 10, 1453-65.

TROMPOUKI, E., BOWMAN, T. V., LAWTON, L. N., FAN, Z. P., WU, D. C., DIBIASE, A., MARTIN, C. S., CECH, J. N., SESSA, A. K., LEBLANC, J. L., LI, P., DURAND, E. M., MOSIMANN, C., HEFFNER, G. C., DALEY, G. Q., PAULSON, R. F., YOUNG, R. A. & ZON, L. I. 2011. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell, 147, 577-89.

TSAI, F. Y. & ORKIN, S. H. 1997. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood, 89, 3636-43.

UCAR, D., BEYER, A., PARTHASARATHY, S. & WORKMAN, C. T. 2009. Predicting functionality of protein-DNA interactions by integrating diverse evidence. Bioinformatics, 25, i137-44.

VAKOC, C. R., MANDAT, S. A., OLENCHOCK, B. A. & BLOBEL, G. A. 2005. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell, 19, 381-91.

VAN HEERINGEN, S. J. & VEENSTRA, G. J. 2011. GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. Bioinformatics, 27, 270-1.

VEENSTRA, G. J., BEUMER, T. L., PETERSON-MADURO, J., STEGEMAN, B. I., KARG, H. A., VAN DER VLIET, P. C. & DESTREE, O. H. 1995. Dynamic and differential Oct-1 expression during early Xenopus embryogenesis: persistence of Oct-1 protein following down-regulation of the RNA. Mech Dev, 50, 103-17.

VELOSO, A., KIRKCONNELL, K. S., MAGNUSON, B., BIEWEN, B., PAULSEN, M. T., WILSON, T. E. & LJUNGMAN, M. 2014. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res, 24, 896-905.

VILLAR, D., BERTHELOT, C., ALDRIDGE, S., RAYNER, T. F., LUKK, M., PIGNATELLI, M., PARK, T. J., DEAVILLE, R., ERICHSEN, J. T., JASINSKA, A. J., TURNER, J. M., BERTELSEN, M. F., MURCHISON, E. P., FLICEK, P. & ODOM, D. T. 2015. Enhancer evolution across 20 mammalian species. Cell, 160, 554-66.

VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009a. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature, 457, 854-8.

VISEL, A., RUBIN, E. M. & PENNACCHIO, L. A. 2009b. Genomic views of distant-acting enhancers. Nature, 461, 199-205.

VOKES, S. A., JI, H., WONG, W. H. & MCMAHON, A. P. 2008. A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. Genes Dev, 22, 2651-63.

WA MAINA, C., HONKELA, A., MATARESE, F., GROTE, K., STUNNENBERG, H. G., REID, G., LAWRENCE, N. D. & RATTRAY, M. 2014. Inference of RNA polymerase II transcription dynamics from chromatin immunoprecipitation time course data. PLoS Comput Biol, 10, e1003598.

WANG, H., CAO, R., XIA, L., ERDJUMENT-BROMAGE, H., BORCHERS, C., TEMPST, P. & ZHANG, Y. 2001. Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. Mol Cell, 8, 1207-17.

WANG, H. G., MORAN, E. & YACIUK, P. 1995. E1A promotes association between p300 and pRB in multimeric complexes required for normal biological activity. J Virol, 69, 7917-24.

WANG, Q., CARROLL, J. S. & BROWN, M. 2005. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. Mol Cell, 19, 631-42.

WANG, Z., ZANG, C., CUI, K., SCHONES, D. E., BARSKI, A., PENG, W. & ZHAO, K. 2009. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell, 138, 1019-31.

WANG, Z., ZANG, C., ROSENFELD, J. A., SCHONES, D. E., BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., PENG, W., ZHANG, M. Q. & ZHAO, K. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet, 40, 897-903.

WASYLYK, B., WASYLYK, C., AUGEREAU, P. & CHAMBON, P. 1983. The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell, 32, 503-14.

WATANABE, M. & WHITMAN, M. 1999. FAST-1 is a key maternal effector of mesoderm inducers in the early Xenopus embryo. Development, 126, 5621-34.

WEBER, H., SYMES, C. E., WALMSLEY, M. E., RODAWAY, A. R. & PATIENT, R. K. 2000. A role for GATA5 in Xenopus endoderm specification. Development, 127, 4345-60.

WEIRAUCH, M. T., YANG, A., ALBU, M., COTE, A. G., MONTENEGRO-MONTERO, A., DREWE, P., NAJAFABADI, H. S., LAMBERT, S. A., MANN, I., COOK, K., ZHENG, H., GOITY, A., VAN BAKEL, H., LOZANO, J. C., GALLI, M., LEWSEY, M. G., HUANG, E., MUKHERJEE, T., CHEN, X., REECE-HOYES, J. S., GOVINDARAJAN, S., SHAULSKY, G., WALHOUT, A. J. M., BOUGET, F. Y., RATSCH, G., LARRONDO, L. F., ECKER, J. R. & HUGHES, T. R. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. Cell, 158, 1431-1443.

WEST, A. G., GASZNER, M. & FELSENFELD, G. 2002. Insulators: many functions, many mechanisms. Genes Dev, 16, 271-88.

WHEELER, G. N. & BRANDLI, A. W. 2009. Simple vertebrate models for chemical genetics and drug discovery screens: lessons from zebrafish and Xenopus. Dev Dyn, 238, 1287-308.

WIJGERDE, M., GROSVELD, F. & FRASER, P. 1995. Transcription complex stability and chromatin dynamics in vivo. Nature, 377, 209-13.

WILSON, P. A. & HEMMATI-BRIVANLOU, A. 1995. Induction of epidermis and inhibition of neural fate by Bmp-4. Nature, 376, 331-3.

WOLF, D., RODOVA, M., MISKA, E. A., CALVET, J. P. & KOUZARIDES, T. 2002. Acetylation of beta-catenin by CREB-binding protein (CBP). J Biol Chem, 277, 25562-7.

WOLPERT, L. 2011. Principles of Development, New York, USA, Oxford University Press.

WU, C. 1980. The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature, 286, 854-60.

WU, J., HUANG, B., CHEN, H., YIN, Q., LIU, Y., XIANG, Y., ZHANG, B., LIU, B., WANG, Q., XIA, W., LI, W., LI, Y., MA, J., PENG, X., ZHENG, H., MING, J., ZHANG, W., ZHANG, J., TIAN, G., XU, F., CHANG, Z., NA, J., YANG, X. & XIE, W. 2016. The landscape of accessible chromatin in mammalian preimplantation embryos. Nature, 534, 652-7.

WU, W. S. & LAI, F. J. 2015. Functional redundancy of transcription factors explains why most binding targets of a transcription factor are not affected when the transcription factor is knocked out. BMC Syst Biol, 9 Suppl 6, S2.

WUHR, M., FREEMAN, R. M., JR., PRESLER, M., HORB, M. E., PESHKIN, L., GYGI, S. & KIRSCHNER, M. W. 2014. Deep proteomics of the Xenopus laevis egg using an mRNA-derived reference database. Curr Biol, 24, 1467-1475.

XI, H., SHULHA, H. P., LIN, J. M., VALES, T. R., FU, Y., BODINE, D. M., MCKAY, R. D., CHENOWETH, J. G., TESAR, P. J., FUREY, T. S., REN, B., WENG, Z. & CRAWFORD, G. E. 2007. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. PLoS Genet, 3, e136.

XIAO, T., HALL, H., KIZER, K. O., SHIBATA, Y., HALL, M. C., BORCHERS, C. H. & STRAHL, B. D. 2003. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev, 17, 654-63.

YACIUK, P. & MORAN, E. 1991. Analysis with specific polyclonal antiserum indicates that the E1A-associated 300-kDa product is a stable nuclear phosphoprotein that undergoes cell cycle phase-specific modification. Mol Cell Biol, 11, 5389-97.

YANG, J., TAN, C., DARKEN, R. S., WILSON, P. A. & KLEIN, P. S. 2002. Beta-catenin/Tcf-regulated transcription prior to the midblastula transition. Development, 129, 5743-52.

YAO, T. P., OH, S. P., FUCHS, M., ZHOU, N. D., CH'NG, L. E., NEWSOME, D., BRONSON, R. T., LI, E., LIVINGSTON, D. M. & ECKNER, R. 1998. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell, 93, 361-72.

YOUNGER, S. T. & RINN, J. L. 2017. p53 regulates enhancer accessibility and activity in response to DNA damage. Nucleic Acids Res, 45, 9889-9900.

ZEITLINGER, J., SIMON, I., HARBISON, C. T., HANNETT, N. M., VOLKERT, T. L., FINK, G. R. & YOUNG, R. A. 2003. Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. Cell, 113, 395-404.

ZHANG, C., BASTA, T., FAWCETT, S. R. & KLYMKOWSKY, M. W. 2005. SOX7 is an immediate-early target of VegT and regulates Nodal-related gene expression in Xenopus. Dev Biol, 278, 526-41.

ZHANG, C., BASTA, T., JENSEN, E. D. & KLYMKOWSKY, M. W. 2003. The beta-catenin/VegT-regulated early zygotic gene Xnr5 is a direct target of SOX3 regulation. Development, 130, 5609-24.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol, 9, R137.

ZHOU, H., KAPLAN, T., LI, Y., GRUBISIC, I., ZHANG, Z., WANG, P. J., EISEN, M. B. & TJIAN, R. 2013. Dual functions of TAF7L in adipocyte differentiation. Elife, 2, e00170.

ZHU, H., DOHERTY, J. R., KULIYEV, E. & MEAD, P. E. 2009. CDK9/cyclin complexes modulate endoderm induction by direct interaction with Mix.3/mixer. Dev Dyn, 238, 1346-57.

ZON, L. I. 1995. Developmental biology of hematopoiesis. Blood, 86, 2876-91.